From Department of Laboratory Medicine
Karolinska Institutet, Stockholm, Sweden

INFLUENCE ON THE TRANSCRIPTOME OF TEC FAMILY KINASES WITH SPECIAL EMPHASIS ON BTK

Hossain M. Nawaz

Karolinska Institutet

Stockholm 2013

Published by Karolinska Institutet. Printed by E-print AB 2013

ISBN 978-91-7549-311-4

Science is a way of thinking much more than it is a body of knowledge.

Carl Sagan (1934-1996), American Scientist

To my family

ABSTRACT

Over the last decade, scientists all over the world have profoundly used gene expression profiling based on microarray. Affymetrix is considered as the one of the pioneer platforms in the field of microarray technology. In this thesis, the Affymetrix Genechip® arrays were used to study the transcriptome of Tec family kinases with special emphasis on Bruton's tyrosine kinase (Btk) in avian B-lymphoma DT40 cell-line and fruit flies (*Drosophila melanogaster*). Btk is a protein tyrosine kinase belonging to the Tec family of kinases (TFKs). Btk is involved in signal transduction of the B cell receptor (BCR) pathway and plays an essential role in B lymphocyte development and function. X-linked agammaglobulinemia (XLA) is a primary immunodeficiency disease caused by mutations in the *BTK* gene. We studied Btk-deficient DT40 avian cell line reconstituted with the human *BTK* gene in order to investigate whether the loss-of-function can be rescued by the gene substitution at the transcriptomic level. Differences in the gene expression pattern showed statistically significant changes between parental DT40 and all the Btk KO cell populations, irrespective of whether they are reconstituted or not. Our result showed clonal selection of Btk knockout and gene reconstituted cells.

Btk is highly conserved during evolution appearing in sponges and in *Drosophila melanogaster*. In Drosophila, *Btk29A* (*Btk family kinase at 29A*) is the sole kinase that represents the Tec family, and it is most similar to Btk itself in terms of overall homology. In fact, the protein product from the type 2 splice variant of this gene exhibits the highest homology to Btk among the five mammalian TFKs. The type 1 splice variant has a shorter N-terminus that is unique to *Drosophila* Btk29A. *Btk29A* displays a dynamic pattern of expression through the embryonic to adult stages. $Btk29A^{ficP}$ is a unique allele in that it is devoid of the type 2 isoform while leaving type

1 isoform intact. The *Btk29A*^{ficP} mutants survive to the adult stage, exhibiting a copulation defect and reduced lifespan after eclosion. We compared Btk mutant flies with their revertant strains using microarray gene expression profiling in adult brain and larvae CNS in order to investigate whether the loss-of-function phenotype can be rescued at the transcriptomic level. The whole transcriptomic profile for the different sample groups revealed Gene Ontology patterns for lifespan abnormalities in adult head neuronal tissue, but not in larval stage. We also carried out cross-species comparison in Btk-deficient flies and mice, which showed no significant overlap of the transcriptomic changes. Our results suggest that the evolutionary conservation is confined to components of the proximal signaling, whereas the corresponding, downstream transcriptional regulation has been differentially wired.

In paper III (manuscript), we extended a previous study addressing the influence of another TFK, namely Itk, by investigating the influence of co-culture the interaction of $CD4^+$ and $CD8^+$ T cells, when they are separated or not in the context of T cell activation using expression profiling data. T cell activation is one of the important steps in the immune response where hundreds of genes and proteins play crucial roles in a highly organized manner. Activation of those genes in the immune system is dependent on the T cell receptor signaling pathway and its regulation of transcription. Our result shows 6% of the transcripts are influenced by the contact between $CD4^+$ and $CD8^+$ T cells. This finding is of general importance, since whenever T cells are cultivated, whether they are separated into subsets or not, influences the analysis of their transcriptomes.

LIST OF PUBLICATIONS

I. **Hossain M Nawaz**, K Emelie M Blomberg, Jessica M Lindvall, Tomohiro Kurosaki, C I Edvard Smith.
Expression profiling of chicken DT40 lymphoma cells indicates clonal selection of knockout and gene reconstituted cells. Biochemical and Biophysical Research Communications 10/2008; 377(2):584-8.

II. **Hossain M Nawaz**, Per Kylsten, Noriko Hamada, Daisuke Yamamoto, C I Edvard Smith, Jessica M Lindvall
Differential evolutionary wiring of the tyrosine kinase Btk.
PLoS ONE . 01/2012; 7(5):e35640.

III. **Hossain M Nawaz**, K Emelie M Blomberg, Jessica M Lindvall, C I Edvard Smith and Anna Berglöf
Estimating the influence of CD4+ CD8+ T cell interactions on the transcriptome (Manuscript).

CONTENTS

LIST OF ABBREVIATIONS

APC	Antigen presenting cell
BAFF-R	B-cell activator of the TNF-a family –receptor
BCR	B cell receptor
BLNK	B-cell linker protein
BM	Bone marrow
BMX	Bone marrow tyrosine kinase on chromosome X
Btk	Bruton's tyrosine kinase
CD	Cluster of differentiation
cDNA	Complementary DNA
DAG	Diacylglycerol
dChip	DNA-chip analyzer
EBF1	Early B cell factor
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FDR	False discovery rate
GC	Germinal center
GCOS	Affymetrix Genechip® operating system
GEO	Gene expression omnibus
GSEA	Gene set expression analysis
GO	Gene ontology
GPCR	G-protein-coupled receptors
IFN-γ	Interferon-γ
Ig	Immunoglobulin
IL	Interleukin
IP_3	Inositol (1,4,5) triphospatase
IRAK	Interleukin-1 receptor (IL-1R)-associated kinase-1
ITAM	Immunoreceptor tyrosine-based activation motif
Itk	IL-2 inducible T cell kinase
IVT	In vitro transcription
JNK	c-Jun N-terminal kinase
Lat	Linker for activation of T cells
Lck	Lymphocyte-specific protein tyrosine kinase
MAPK	Mitogen-associated protein kinase
Mal	MyD88 adapter-like protein
MCHC	Major Histocompatibility Complex
MoAb	Monoclonal antibody
MM	Mismatch
Myd88	Myeloid differentiation protein 88
MZ	Marginal Zone
NGS	Next-generation sequencing
NFAT	Nuclear factor of activated T cell
NF- κB	Nuclear factor kappa B
PCA	Principal component analysis
PCR	Polymerase chain reaction

PH	Plectstrin homology
PI3K	Phosphoinositide 3 kinase
PIP_2	Phosphatidylinositol-4,5-bisphosphonate
PIP_3	Phosphatidylinositol-3,4,5-triphosphonate
PKB	Protien kinase B
PLCγ	Phspholipase C gamma
PM	Perfect match
Pre-B-cell	Precursor B cell
Rlk/Txk	Resting lymphocyte kinase/T and X cell expressed kinase
RMA	Robust Multi-array average
Sab	SH3-domain binding protein
SH	Src homology
STAT	Signal tranducer and activator of transcription
Syk	Spleen tyrosine kinase
Tc	Cytotoxic T cell
TCR	T-cell receptor
Tec	Tyrosine kinase expressed in hepatocellular carcinoma
TFKs	Tec family kinases
TH	Tec homology
Th	T helper cell
TNF-β	Tumor necrosis factor β
TLR	Toll like receptor
Xid	X-linked immunodeficiency
XLA	X-linked agammaglubulinemia
Zap70	Zeta-chain-associated protein kinase 70

1 INTRODUCTION

Recent development in bioinformatics and high-throughput technologies such as microarray analysis has made a revolution in the field of molecular medicine. This has a great impact on the discovery of new targets for the treatment of disease in form of drug development, immunotherapeutic and gene therapy. Microarray studies and other genomic techniques have helped the scientists to understand mechanisms underlying normal and dysfunctional biological processes such as cancer, inflammation and immunodeficiency. This has now become an emerging discipline that integrates high-throughput expression profiling technology and systems biology approaches [1]. With the rapid advances of microarray analysis technologies, bioinformatics is extensively used in the studies of gene expression profiling, miRNA–mRNA interactions, DNA–protein interactions and protein–protein interactions. Different microarray platforms have been developed so far is based on same principle: hybridization between two complementary strands of nucleic acid. Purified RNA is fluorescently- or radioactively labeled and hybridized to the slide/membrane. After thorough washing, laser scanning obtains the raw data; the data may then be entered into a database and analyzed by a number of statistical methods with help of different bioinformatics tools [2].

In this thesis work, I have used microarray expression profiling in order to investigate transcriptional regulation in Tec family kinases with a special emphasis on Bruton's tyrosine kinase (Btk) in chicken DT40 cell line as well as in fruit flies and transcriptional influence in activated primary mouse T cells.

1.1 Tec family kinases

Tec family kinases (TFKs) form the second largest family of cytoplasmic tyrosine kinases in mammalian cells and include, in addition to Btk, Tec (Tyrosine kinase expressed in hepatocellular carcinoma), Itk (IL-2 inducible T cell kinase), Txk (Tyrosine protein kinase , also known as Rlk), and Bmx (bone marrow tyrosine kinase gene on chromosome X) [3]. The name tyrosine kinase stems from the ability of these enzymes to phosphorylate tyrosine residues. The family has been so far described in mammals and semantics even in the unicellular organism *Monosiga brevicollis* [4]. Most of the TFKs (Btk, Itk, Tec, Txk) are predominantly expressed in hematopoietic cells mostly in B and T lymphocytes, whereas Bmx is essentially restricted to endothelial cells [5-8]. In B cells, Btk and Tec are highly expressed; whereas T cells mostly express Itk, Tec and Rlk/Txk [9]. TFKs have been also showed to be expressed in the myeloid lineage in macrophages (Bmx, Btk and Tec), in neutrophils (Bmx, Btk and Tec), in mast cells (Btk, Itk and Tec) and in dendritic cells (Btk) [10]. These kinases are involved in a wide range of signaling pathways that control mitogen-activated protein kinase (MAPK) activation, Ca^{2+} influx, actin reorganization, transcriptional regulation, cell survival and cellular transformation [11-15]

Our group has previously shown that NF-kappaB (NF-κB) regulates the transcription Tec and identified a unique functionally active NF-κB binding site in its promoter [16]. Txk has also been shown to and regulate interferon-γ gene transcription in Th1 cells by forming a complex with poly (ADP-ribose) polymerase 1 and elongation factor 1α [17]. Itk has also been reported to interact physically with transcription factor TFII-I and potentiates TFII-I-driven c-fos transcription [18]. Role of Btk in transcriptional regulation is described in more details in later part.

1.2 Bruton's tyrosine kinase (Btk)

Expression of Btk has been detected at all stages of B-lymphocytes prior to the plasma cell stage [9]. Btk is also involved in mast cell activation [19]. It has been recently reported that Btk expression is up-regulated during maturation and activation of mouse Natural Killer (NK) cells [20] and is also important for effective osteoclastogenesis as its deficiency has resulted in incomplete osteoclast differentiation and mild osteopetrosis [21]. Btk is also known to play a dual role when it comes to survival and apoptosis [22]. Btk is highly conserved during evolution appearing in sponges and in *Drosophila melanogaster* and vertebrates [4]. A sign of the conservation is the fact that the human *BTK* gene can partly substitute for loss-of-function of the *D. melanogaster Btk29A gene* [23].

X-linked agammaglobulinemia (XLA) is a primary immunodeficiency disease caused by mutations in the *BTK* gene [9, 24-26], that was first described by Dr. Ogden Bruton in 1952. Patients with XLA lack circulating B cells due to a complete block in their B lymphocyte development at the mature B cell stage[27, 28]. Therefore, patients have severe hypogammaglobulinemia and impaired antibody responses. This results in increased susceptibility to bacterial and enteroviral infections. The X-linked immunodeficiency (*Xid)* phenotype in mice is caused by a missense mutation (R28C) in the N-terminally located pleckstrin homology (PH) domain of *Btk*. Btk knockout (KO) mice, which lack detectable Btk protein, have an *Xid*-like phenotype [29]. Recent investigations also suggest the role of Btk in the development and BCR/TLR-dependent activation of AM14 rheumatoid factor B cells and Btk deficiency only partially blocks the development of systemic autoimmune disease [30].

1.3 *BTK* gene and transcription factors

Btk gene was first cloned in 1992 in two different research groups in Europe and USA [24, 25]. The *BTK* gene is located on the human X chromosome at Xq21.3-Xq22 and consists of 19 exons spanning 37.5 kb of genomic DNA [31]. It has been reported that Btk may play a role in transcription in addition to its involvement in the signaling cascade from the BCR. Several laboratories, including ours [32], have shown that Btk affects downstream components of the BCR signaling pathway, thereby regulating gene expression. As an example, Btk has been shown to interact with several transcription factors, including nuclear factor of activated T cell (NFAT) and NF-κB [16, 33] implicating that it has a direct role in transcriptional regulation [34]. A review article from Wasif A. Khan, depicted that *BTK* is required for efficient induction of bcl-x transcription in response to BCR crosslinking which is regulated by NF-κB [35].

Several other research groups have reported the role of *BTK* in transcriptional regulation. Among those are the transcription factor BAP-135/TFII-I [36] and Bright [33, 37]. Recently, it has been demonstrated that in addition to Btk, Bright can associate with TFII-I to induce immunoglobulin heavy chain transcription [34, 38]. Btk has also been shown to induce the transcriptional co-activation activity of the Btk-associated protein BAM 11[39]. Finally, Btk has been shown to phosphorylate signal transducer and activator of transcription 5A (STAT5A) in the chicken B-cell line DT-40 [40].

1.4 Btk family kinase at 29A (Tec29) in Drosophila

In Drosophila, *Btk29A* (*Btk family kinase at 29A*) is the sole kinase that represents the Tec family, and it is most similar to Btk itself in terms of overall homology. *Btk29A* is required for survival and male genital formation in Drosophila. *fickle*^{P} (*fic*^{P}), which removes the Btk homolog in Drosophila, was isolated in a screening of P-element mutants with defects in mating behavior [41]. *Btk29A* has splice variants (type 1 and type 2), generated by an alternative exon usage (Figure 1). Both the transcripts are expressed in the central nervous system (CNS) and imaginal discs; the domains and/or timing of expression in these tissues are distinct from each other. Complete loss-of-function of the gene (i.e., loss of both types 1 and 2) causes oocyte undergrowth and embryonic death accompanied by defective head involution [42-44], while selective loss of the type 2 transcripts spares life but reduces the life span in the adult and leads to malformation of the male genitalia [23]. It has been also reported that Btk-binding protein Sab (SH3-domain binding protein), a Src homology 3 domain-binding protein inhibits the auto- and transphosphorylation activity of Btk and thus functions as a transregulator of Btk [45].

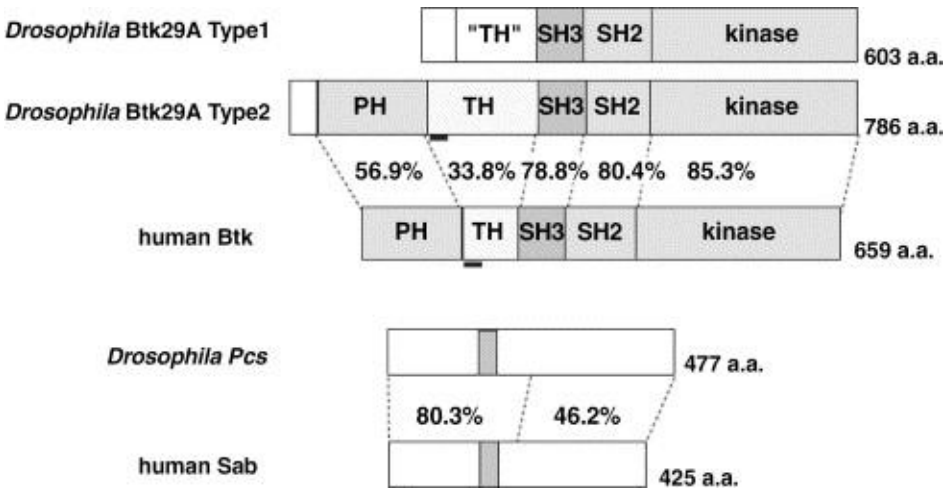

Figure 1: Structural comparison of Drosophila and human orthologs of Btk. Reproduced from Baba et al., 1999 [41].

1.5 B cell receptor signaling

The B cell antigen receptor (BCR) is composed of a transmembrane immunoglobulin heavy chain (IgH) and a covalently associated light chain (IgL) with two heterodimers Ig-α (CD79a) and Ig-β (CD79b) [46]. Ig-α and Ig-β each contain a unique 26 amino acid region immunoreceptor tyrosine-based activation motif (ITAM) in their cytoplasmic terminal domains which functions as a docking site for many of the signal transducing modular proteins during BCR cross-linking. This regulates the antigen recognition component and the signal transduction by BCR, which in turn modulates gene expression, adhesion or survival, thereby determining the fate of antigen-encountered B cells [47].

The initial signaling pathways activated in response to BCR ligation involve multiple protein tyrosine kinase cascades such as Src family kinase Lyn, spleen tyrosine kinase (Syk), Bruton's tyrosine kinase (Btk) and adaptor protein B-cell linker (BLNK) [48, 49]. Btk is first phosphorylated by Lyn, at tyrosine 551[50], that leads to the activation of phospholipase C (PLC)-γ2 that catalyzes the production of second messengers by hydrolysis of phosphatidylinositol-4,5-bisphosphate [PtdIns(4,5)P2] to diacylglycerol (DAG) and inositol-1,4,5-trisphosphate ($IP_{3)}$. IP3 binds to its receptor (IP3R) in the endoplasmic reticulum (ER) membrane, thereby causing a rapid release of Ca^{2+} from ER stores [51]. These second messengers regulate calcium mobilization and activate the PKC isoenzymes that are essential for functional expression of transcription factors.

1.6 T cell biology and activation

T cells are key effectors of the adaptive immune response as they play important roles in the immune defense and are major effectors in autoimmune diseases. T cell activation takes place upon ligation of T cell receptors (TCRs) by MHC molecules on antigen-presenting cells (APCs) (Figure 2). Itk, one of the family members of non-receptor tyrosine kinases plays a crucial role for TCR –induced activation and IL-2 production [12, 52]. Following antigen-receptor engagement in T cells, Itk induces tyrosine phosphorylation of PLCγ2 with subsequent production of IP_3 and mobilization of intracellular Ca^{2+} from ER. This in turn activates calmodulin and Ca^{2+} sensitive phosphatase Calcineurin and regulates transcription factors such as NFAT [53].

The SRC family kinase (SFK) members Lck and Fyn are the first molecules to be activated following TCR engagement [54-56]. SFKs are essential to provide the tonic signaling that is required for the survival of naïve T cells [57]. Lck is constitutively active in naïve T cells and maintains a basal level of phosphorylation on TCR-associated ζ-chains [58]. Lck binds to the cytoplasmic domains of the TCR co-receptors CD4 and CD8 [59]. Following interaction of the TCR with a peptide–MHC complex, these co-receptors are important for targeting the delivery of Lck into close proximity to its substrates: TCR-associated CD3 and ζ-chain immunoreceptor tyrosine-based activation motifs (ITAMs) [60]. These targets of Lck including the tyrosine residues in the ITAMs of TCR-associated CD3γ chain, CD3δ chain, CD3ε chains and the ζ-chains, and the SYK family kinase Zap70 (ζ-chain associated protein kinase of 70 kDa) get phosphorylated TCR[61]. This TCR 'signal triggering module' involves Lck-dependent phosphorylation of the ITAMs, allowing the recruitment and the Lck-dependent phosphorylation of Zap70 [62]. This aggregation and phosphorylation

results in conformational changes in Zap70, which promote its kinase activity, leading to the phosphorylation of its target molecules; including the key adaptor molecule linker for activation of T cells (Lat) [63]. In turn, the phosphorylated tyrosine residues of Lat recruit multiple downstream adaptor and signaling molecules, which then activate several signaling cascades and exert cellular functions like cell adhesion; mobilization to the nucleus of transcription factors that are crucial for the expression of genes necessary for T cell growth and differentiation; and actin reorganization, which is essential for T cell activation, proliferation and adhesion.

Itk is also important in signaling downstream of the T cell co-stimulatory receptor CD28 [64, 65]. CD28 is activated by B7-1 and B7-2 [66] and activates the Ser/Thr kinase protein kinase B which regulates the PI3K pathway [67]. The major signaling product of PI3K activity is phosphatidylinositol (3,4,5)-trisphosphate [PI(3,4,5)P_3], which rapidly accumulates within cells following cellular stimulation. This molecule is largely responsible for the co-localization of 3′-phosphoinositide-dependent protein kinase (PDK)-1 and its substrate protein kinase B (PKB)–Akt at the plasma membrane, through interaction with their pleckstrin homology (PH) domains. PDK-1 phosphorylates and activates the AGC kinase members including PKB–Akt, which is a key effector of PI3K signaling that is sufficient to co-stimulate interleukin (IL)-2 and interferon-γ production. Optimal T-cell activation requires T-cell receptor (TCR) signaling as well CD28 T-cell co-stimulatory molecule also known as ‘two-signal’ model of T-cell activation [68, 69].

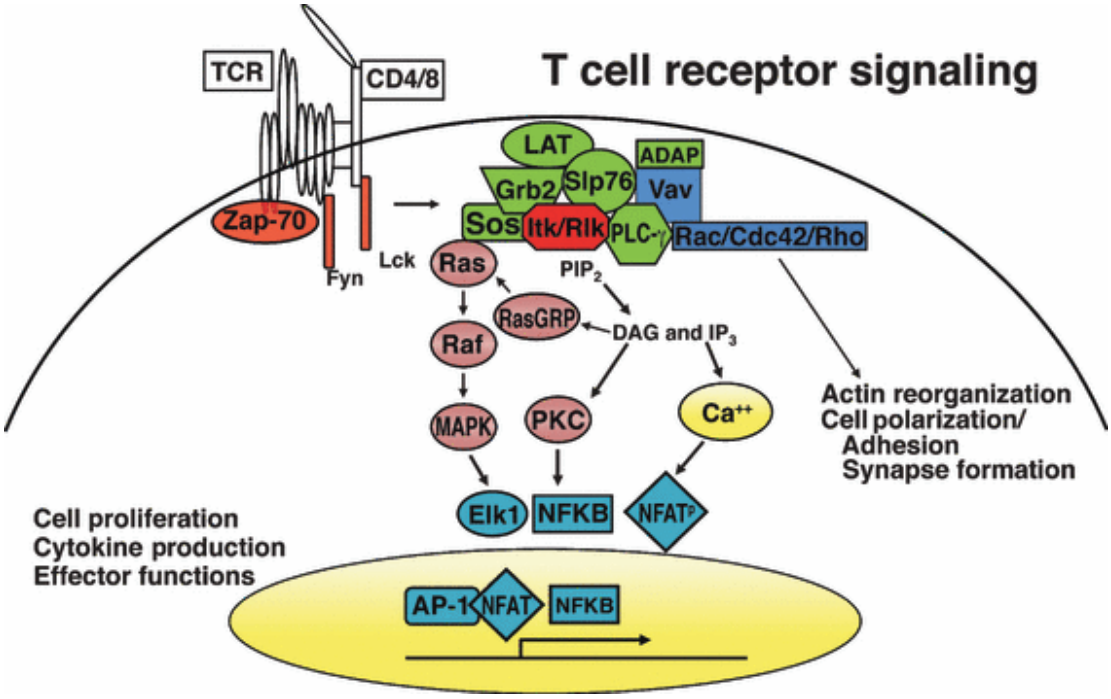

Figure 2: T cell receptor signaling. Reproduced from Readinger et al., 2009 [70].

1.6.1 CD4$^+$ and CD8$^+$ T cells

CD4 and CD8 T cell surface molecules play a role in T cell recognition and activation by binding to their respective class II and class I major histocompatibility complex (MHC) ligands on an antigen presenting cell (APC). Naïve CD4$^+$ and CD8$^+$ T cells get activated in the peripheral lymphoid organs after receiving external signals or stimuli in order to proliferate and differentiate into effector and regulatory cells [71-73]. *In vitro* stimulation of T lymphocytes has been frequently used to expand specific cells present at low precursor frequency *in vivo*[74]. However, this technique might not reflect the *in vivo* situation because of the variable efficiency of outgrowth of different T cell sub-populations [75]. Interaction of the CD4$^+$/CD8$^+$ lineages in the thymus is a complex process between these functional subsets[76]and their surface-phenotype is crucial for the entire immune response [77].

1.7 Global gene expression profiling

Gene expression technology is currently a widely used and powerful tool of molecular biology for investigating the transcriptional behavior of biological systems as well as for classifying cell states in disease. There are to date over 750,000 gene expression data sets in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) public database [78]. Global gene expression analysis uses DNA microarrays, RNA-Seq, and other methods to measure the levels of RNA species in biological systems [79-82]. Among those, DNA microarrays have been most frequently used for expression analysis and it consist of millions of individual oligonucleotide probes fixed to a solid surface. These oligonucleotide probes have sequences representative of known RNA species for the whole genome or a specific part of the genome and are generally used to quantitate the relative levels of RNA species that hybridize to the probes. More recently developed sequencing technologies known as Next-generation sequencing (NGS) also provide a measure of the frequency of RNA species through sequencing of RNA-derived cDNA populations [83-85]. There are also other approaches that have been developed in last few years such as digital molecular barcoding, which represents a fusion of the hybridization and counting approaches [86]. One mostly used platform is the Nanostring nCounter assay system which digitally quantifies the hybridization of labeled probes to RNA molecules and single-molecule imaging to provide a measurement of the frequency of particular RNA species [87].

1.7.1 Microarray expression profiling

The very first article describing microarrays was published in 1995 by Brown and co-workers [88] where they showed differential expression measurements of 45 *Arabidopsis thaliana* genes using cDNA arrays. Previously described DNA

microarrays are small, solid supports onto which the sequences from thousands of different genes are immobilized, or attached, at fixed locations [89]. These are usually made by glass microscope slides, silicon chips, or nylon membranes where the DNA is printed or spotted [90]. Each spot contains one type of nucleic acid and each nucleic acid is a polymer of nucleotides of defined length. The nucleic acid on each spot differ from the other nucleic acids in the array in the sequence of the bases adenine (A), cytosin (C), guanine (G), and thymine (T) [91, 92]. DNA arrays are based on the specific base pairing of complementary nucleotides (A-T and C-G) leading to double-stranded sequences of nucleic acids. In cDNA microarrays, probes are usually with a length of 400–1,000 base pairs (bp), which leads to a high specificity for the targets. There are basically three types of samples that can be used to construct DNA microarrays, two are genomic and the other is "transcriptomic", that is, it measures mRNA levels. The target DNA used will also determine the type of control and sample DNA that is used in the hybridization solution [93, 94].

The oligonucleotides microarrays are more popular array platforms at present where the probes are with lengths <200 bp, up to short oligonucleotides of ~18-25 bases in case chips [95]. There are several commercially available arrays in the market and among those, Affymetrix, Agilent, Illumina and NimbleGen are the mostly used microarray platforms. Affymetrix has 25mer probes and are synthesized by Photolithographic technique [96, 97]. Agilent uses 60mer probes and synthesized by inject technology on glass slides [98]. Illumina has 50mer probes and they use bead technology [99]. NimbleGen uses 60mer probes that are synthesized on a glass slide [100] .

1.7.2 Affymetrix Genechip® technology

One of the earliest descriptions of today's most frequently used platform for expression profiling, Affymetrix GeneChips, is from 1996 [101] and many advances have been made over the last decade. Affymetrix Genechip technology brought revolution in the field of bio-medical research with their different types of arrays (Figure 3). Affymetrix provide two generations of expression arrays, 3' (IVT) and 1.0 ST. 3' arrays were the first array type from Affymetrix, and are available for a large number of species. Classical 3' expression microarrays are not able to discriminate between alternatively spliced transcripts that have identical 3' ends. The combination of Affymetrix' Whole Transcript (WT) Assay and high-density arrays, including GeneChip® Exon 1.0 ST Arrays and GeneChip® Gene 1.0 ST Arrays, provides a more complete and more accurate picture of overall gene expression. With approximately four probes per exon and roughly 40 probes per gene, the GeneChip® Human Exon 1.0 ST Array enables two complementary levels of analysis—gene expression and alternative splicing. The most advance is to date is the GeneChip® miRNA 2.0 array covering 71 organisms, including human, mouse, rat, and monkey.

Affymetrix use 'Photolithographic' technique where large-scale production of GeneChip probe arrays containing hundreds of thousands of oligonucleotide sequences on a glass "chip" about 1.5 cm^2 in size can be possible. Due to their very high information content, GeneChip probe arrays are finding widespread use in the hybridization-based detection and analysis of mutations and polymorphisms ("genotyping"), and in a wide range of gene expression studies [97].

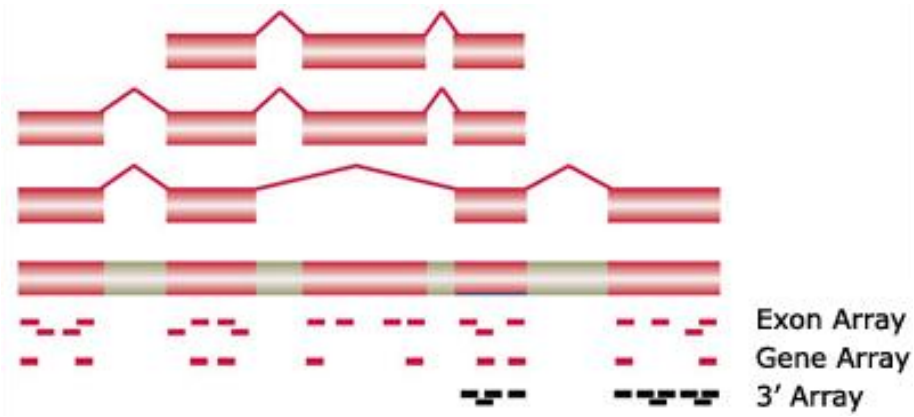

Figure 3: Different types of Affymetrix arrays [102].

Affymetrix conventionally uses several probe-sets for each transcript or genes and there are principally two types of probes: (1) probes that have complete complementarity to their target sequence [perfect match probe (PM)] and (2) probes with a single mismatch to the target, centered in the middle of the oligonucleotide [mismatch probe (MM)] (Figure 4). The PM/MM design is used for identification and subtraction of nonspecific hybridization and background signals[103]. MM probes are effective internal controls, since they will hybridize to nonspecific sequences about as effectively as their counterpart PM probes. This unique probe pairing strategy helps identify and subtract nonspecific hybridization and background signals. The new Exon arrays, Gene1.0 ST arrays, SNP 5.0, 6.0 arrays do not contain MM probes.

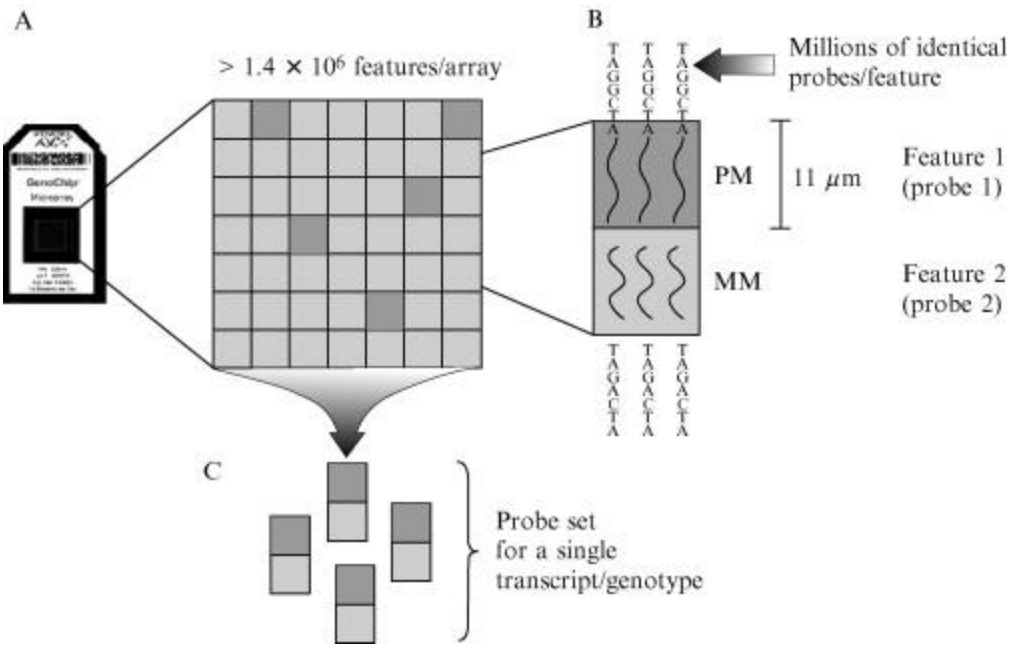

Figure 4: Basic principle of IVT arrays (Re-produced from Dalma-Weiszhausz et al., 2006) [104].

The target preparation protocols have several steps and the first step start with a purified nucleic acid sample, mostly RNA. The RNA is usually amplified and then labeled and fragmented consequently. The labeled RNA is then purified and fragmented by hydrolysis. The most widely used sample preparation for gene expression utilizes the IVT reaction as originally described by Eberwine and colleagues [105]. In this assay cDNA synthesis is initiated from an oligo(dT) primer that is also coupled to a T7 RNA polymerase primer and the cDNA synthesis starts adjacent to the poly(A) tail of the mRNA. After second strand synthesis, a double-stranded cDNA copy of each mRNA is created attached to the T7 RNA polymerase primer. An IVT reaction is then carried out to create a biotinylated RNA target (Figure 5). A variation of this technique utilizes two rounds of IVT amplification and is used to create a target from very small amounts (100 ng or less) of starting material.

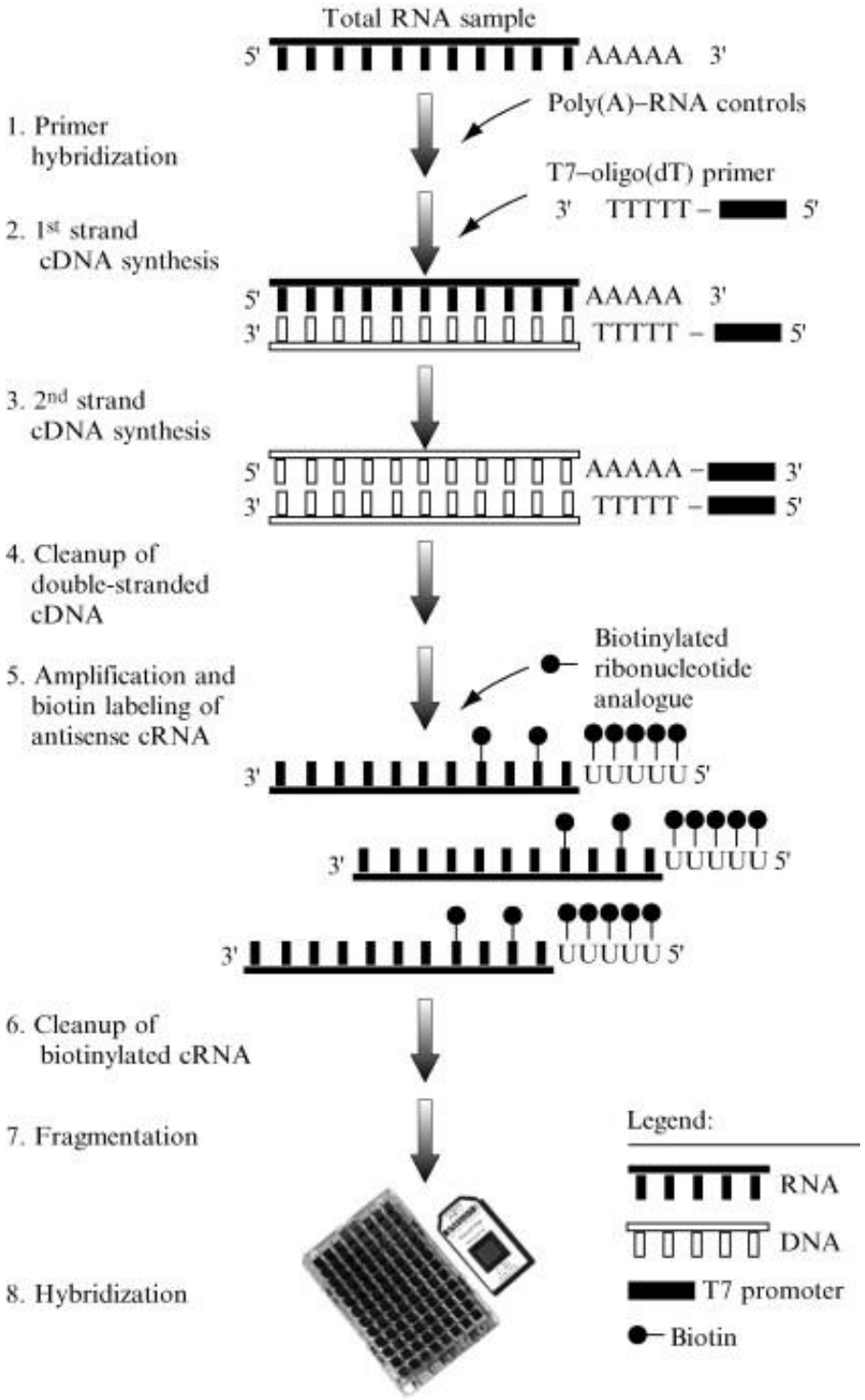

Figure 5: Schematic representation of array hybridization [www.affymetrix.com].

1.7.3 Microarray data analysis

After hybridization and washing, the chip is being scanned and the next step in analysis is the assignment of pixels that make up the image (.dat) file. Several methods have been developed so far and global gridding method is widely used for the Affymetrix array. In this method, the four corners of the array, defined by checkerboard patterns, serve as anchors for the grid. Features are then created by evenly dividing the area defined by the anchored corners into the known number of features for a given array. An additional step called Feature Extraction has been implemented to assign pixels to features in a more robust manner. This enhances the original pixels assigned to a feature that are shifted as a block, a pixel at a time, and calculate the coefficient of variation (CV) of pixel intensities. Following Feature Extraction the intensity of each feature is calculated and stored in a .CEL file [106].

Regardless of application, the feature intensities found in .CEL files are used by analysis software to detect sequence variation or to differentiate gene expression levels of transcripts. During analysis, the use of multiple probes per genotype or gene is combined with standard statistical methods to provide a transparent and robust conversion of probe intensities to biological information.

1.7.4 Pre-processing of microarray data

Affymetrix chip pre-processing is normally a three-step procedure which include estimation of the background signal, possible background correction normalization, and summarizing the separate normalized probe measurements into a single value per probe set [107]. A variety of algorithms has been used to summarize multiple probe

intensities of a transcript in the experimental sample. 'Detection calls' are made by Affymetrix software through an arithmetic calculation of probe pairs within a set designed to detect a specific transcript (GeneChip MAS5.0 and GCOS software) [108]. More widely used summarization methods include Probe Logarithmic Intensity Error estimation (PLIER) [109], Model based expression method in DNA-chip analyzer (dChip) [110] and Robust Multi-array Average (RMA) [111].

PLIER algorithm utilizes experimental data generated across multiple arrays in order to identify and account for observed patterns in probe intensities difference between probes by means of a parameter termed "probe affinity." Probe affinities determine the signal intensity produced at a specific target concentration for a given probe, and are calculated using experimental data across multiple arrays. The Microarray Suite 5.0 (MAS5.0) is the normalization method developed by Affymetrix. It makes use of the MM measurements to compute an Ideal Mismatch value and then corrects the PM values with these Ideal Mismatch values. MAS5.0 then uses a one-step Tukey Biweight algorithm as summarization method [112] . A new software application called GeneChip RNA EXpression Analysis (GREX) has been introduced recently by Affymetrix that features in addition to MAS5.0 and PLIER: quantile normalization, Analysis of variance (ANOVA) for replicate analysis and integrated retrieval of NetAffx™ Analysis Center annotations for faster identification of biological relevance [102] .

dChip normalizes arrays at PM and MM probe level before computing model-based expression levels. It then uses the Invariant Set Normalization method that is normalization of arrays based on an 'invariant set'. The basic principle is that it chooses a subset of PM probes with small within-subset rank difference in the two arrays, to

serve as the basis for fitting a normalization curve [113, 114]. dChip uses an intensity-modeling approach for data analysis and clusters microarray probe sets exhibiting similar expression profiles. In summary, normalization takes place across all the arrays using probes of 'invariant set' and 'Running median' smoothing with a 'Model-based expression' method for PM-MM or '5^{th} percentile of region (PM-only)' as background subtractions.

The Robust Multi-array Average (RMA) method, first introduced by Irizarry and his group, does not use the MM values [115]. The background is instead estimated by convoluting the signal and noise distributions from the PM values. After background correction, RMA performs quantile normalization and then uses median polish as summarization method. The quantile normalization method aims to make the distribution of probe intensities for each array the same for a set of arrays [116]. When applying the median polish algorithm to a matrix, one subtracts the row medians and then the column medians in an iterative manner. After convergence, the row and column medians are zero, and the remaining matrix is subtracted from the original data matrix. GCRMA [117] is a modified RMA procedure and includes the GC content of the probes in the background adjustment step.

1.7.5 Statistical and down-stream analysis

After pre-processing of the microarray data, the next step is to identify groups of genes with specific expression patterns and relations across the samples. Hierarchical clustering and principal components analysis (PCA) are the methods that are widely used [118]. Hierarchical clustering has the advantage that it is simple and the result can be easily visualized [119]. Hierarchical clustering is an agglomerative approach

in which single expression profiles are joined to form groups, which are further joined until the process has been carried to completion, forming a single hierarchical tree. PCA is a statistical technique for determining the key variables in a multidimensional data set that explain the differences in the observations, and can be used to simplify the analysis and visualization of multidimensional data sets [120]. PCA shows visual estimation of the number of clusters represented in the data which resembles whether genes that have similar, correlated patterns of expression [121].

In order to identify genes that may be affected by a treatment or gene mutation, the simplest and most common experimental set-up is to compare two groups: for example, Treatment vs. Control, or Mutant vs. Wild type, mostly known as pair-wise comparison. The most powerful and commonly used statistical test for comparing a single measure on two groups is the t-statistic, where one gets a *p*-value as an end result. A drawback of the t-statistic for microarray datasets is that most experiments have only a few samples in each group, and so the standard error is not very reliable as there is a risk for false positives. In such cases, one can compute the ratios of variances between groups (F-ratios), and comparing to an F-distribution where discrepancies in variance estimates are not too large for many of the genes that are selected as differentially expressed.

Another way of addressing this is the Non-parametric tests such as Wilcoxon signed-rank test that reflects the distribution of the gene expression values in a strictly symmetric fashion. Permutation testing is an approach that is widely applicable and copes with distributions that are far from Normal. Significance Analysis of Microarray (SAM) is s an interface that implements permutation testing [122, 123]. The meaning of a *p*-value from a permutation procedure differs from the meaning of a

model-based *p*-value. The model-based *p*-value is the probability of the test statistic, assuming that the gene levels in both the treatment and control groups follow the model.

As mentioned above, t–test can give false positive values and for that, statistical filtering is used to specify the false discovery rate (FDR) described by Benjamini and Hochberg [124]. The false discovery rate is an estimate of the number of false positives relative to the number of all genes flagged as 'positives' by a procedure. In contrast the (single test) *p*-value is the probability of a false positive on that one test. In the context of multiple tests, it is an estimate of the number of false positives relative to the number of truly unchanged genes (all negatives). The true multi-test *p*-value is the probability of seeing any false positives among all the tests being done. Another classical multiple testing procedures is the family-wise error rate (FWER) using Bonferroni correction has been used frequently for false positive results [125]. FWER measures the probability that the analysis yields any false positive findings but less used now a days due to its limitations being too conservative for the analysis of microarray data [126].

The *q*-value (also known as corrected *p*-value) is the smallest FDR at which a particular gene would just stay on the list of positives. This is not identical to the p-value, which is the smallest false positive rate (FPR) at which the gene appears positive. The *p*-value is much stricter than the *q*-value. Most of the researchers, who compute significance of genes by permutations, are actually computing the *q*-value, rather than the *p*-value [127].

1.7.6 Gene ontology and Gene set enrichment analysis

The next step is the interpretation of microarray data in a biological context. The Gene Ontology Consortium provides a controlled vocabulary to annotate the biological knowledge we have or that is predicted for a given gene. The Gene Ontologies (GOs) are organized as a hierarchy of annotation terms and the top-level ontologies are *molecular function*, *biological process*, and *cellular component* [128]. The most commonly used methods to test for enrichment of a gene set is known as Gene set enrichment analysis (GSEA) for identifying over/under-representative GO term with a *p*-value or false discovery rate (FDR) [129]. The basic principle of this analysis is that a list of genes is selected from a microarray first, for example, by choosing all significantly differentially expressed genes using a cutoff value based on hypergeometric distribution and use either Fisher's exact test or the χ^2 test. The test for enrichment involves counting how many genes in the gene set occur in the list of selected genes and how many occur on the microarray, and then estimating what proportion of genes from the gene set would be expected to appear in randomly selected lists.

2 AIMS

General aim: The general aim of this thesis work was to identify the role of Btk at the transcriptional level using microarray gene expression analysis, whereas the specific aims of this research were:

1. To define the distinct **gene expression signature** in Btk defective cells
2. To perform cross-species studies (**Meta-analysis**)
3. To investigate the influence of co-culture of different T cell subsets and their activation.

3 MATERIALS AND METHODS

3.1 DT40 cell-line

DT40 is an avian leucosis virus-transformed chicken B-lymphocyte line and is widely used as model system for making gene-targeted mutants. Homologous DNA recombination-controlled gene conversion technique is used here for modification of genetic loci in vertebrate somatic cells [130]. The cell-lines used in **Paper I**, were the parental DT40, Btk knockout, B7.10; Btk KO reconstituted with human wild type Btk, B41.13 or kinase-inactive Btk, B46.5 were generated in Dr. T. Kurosaki's laboratory at RIKEN Research Center for Allergy and Immunology in Japan [131]. The cells were cultured in RPMI 1640 Glutamax medium with 10% FBS, 1% chicken serum, Penicillin-Streptomycin, 4 mM glutamine, and 50 µM 2-mercaptoethanol (2-ME) starting in 1 ml serum-free medium for 15 min at 37 °C with 5% CO_2. Stimulations were during that time done with 5 µl supernatant of the hybridoma cell-line M4 (a kind gift from Dr Max D. Cooper, University of Birmingham, AL). After 15 min incubation, 49 ml medium was added and cells were cultivated for 2, 6 or 24 h with, or without, stimuli.

3.2 Primary T cell separation and stimulation

Scientists have successfully separated sub-populations of T cell in last few decades using different methods [132, 133]. With advancement of immunofluorescence and flow cytometry techniques, isolation of T cell has become an easier and routine method [134, 135]. Both positive and negative selection methods have been developed in order to isolate purified T cell sub-sets [136]. In **Paper III**, T-cells were isolated from suspensions of spleen and lymph nodes of C57BL/6 mice. $CD3^+$, $CD4^+$ and $CD8^+$ T

cells were then isolated by negative selection method [137, 138]. The cell suspensions were incubated with the antibodies in PBS supplemented with 2% FCS. Streptavidin beads (BD Pharmingen) were used for negative depletion according to manufacturer's instructions. The purity of the cells was assessed by flow cytometry and was routinely >90% $CD3^+$, >96% $CD4^+$ and >90% $CD8^+$ T-cells. Separated T cells subsets ($CD3^+$, $CD4^+$ and $CD8^+$) were stimulated with anti-CD3 (1 µg/ml) for 24 h in 48-well plates.

3.3 *Drosophila melanogaster*

The fruit fly, *Drosophila melanogaster* , is a well-studied model organism that is investigated extensively in the twentieth century for understanding molecular mechanisms of human diseases as many basic biological, physiological, and neurological properties are conserved between mammals and *D. melanogaster* [139, 140]. Approximately, 75% of human disease-causing genes are believed to have a functional homolog in the fly [141] and has contributed substantially to further investigate the process of cancer genetics and therapeutic drug discovery [142]. In **Paper II**, we used Btk defective fruit flies, *w;Btk29A*^{ficP}*/CyO* and *w1118; Btk29A*$^{(fic\ Exc1-16)}$*/SM1* were generated in the Yamamoto laboratory [143]. Flies were raised on standard medium on a 12:12 h L:D cycle, at 23 °C and at 55% RH. Flies were anaesthetized using CO_2, then immediately dissected under microscope. The tissues dissected were the complete heads, severed at the neck from adult flies and the CNS (developing brain), including the optic lobes from third instar wandering stage larvae. Tissues were collected into RLT buffer, pooled and extracted for RNA.

3.4 RNA isolation

Different commercially available RNA extraction kits have made RNA isolation more easier and time efficient. In this study, total RNA was extracted with RNeasy Mini Kit (Qiagen, Valencia, CA, USA). The RNA amount and the quality were measured primarily with a NanoDrop spectrophotometer (Wilmington, USA) [144] and then in an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) at the Affymetrix core facility before hybridization. Agilent uses an RNA integrity number (RIN) to get a robust and reliable prediction of RNA integrity, thus checking the quality of the RNA [145]. For **Paper I**, five micrograms of total RNA from each of the ten replicates (per sample) were pooled and hybridized to Genechip® Chicken Genome arrays. For **Paper II-III**, 100 ng of total RNA was amplified according to Affyemtrix Small Sample protocol with two rounds of IVT [146].

3.5 Affymetrix Genechip® array

In **Paper I**, Genechip® Chicken Genome array containing 32,773 transcripts corresponding to over 28,000 chicken genes was used. In total, 28 arrays were generated as a pool of 10 different replicates in each. **In Paper II**, *Drosophila* genome 2 expression arrays comprised of 18,880 probe sets, analyzing over 18,500 transcripts were hybridized and read using standard Affymetrix procedures. In total there were 12 arrays which include three replicates for $Btk29A^{ficP}$ and $Btk29A^{(fic\ Exc1-16)}$ sample groups (the larval CNS and adult heads). In **Paper III**, The GeneChip® Mouse Genome 430 2.0 Array covering the Mouse Expression Set 430 for analysis of over 39,000 transcripts on a single array were used. In total 10 arrays were analyzed which are accessible through the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/).

Microarrays were run at the Bioinformatics and Expression Analysis core facility (http://apt.bea.ki.se/index.html) located at Karolinska Institutet, Huddinge (Novum).

3.6 Data analysis and bioinformatics

In **Paper I,** pre-processing of microarray data was carried out using Affymetrix® Expression Console software. Data were normalized across all the arrays using probes of 'Invariant set' and 'Running median' smoothing. Model-based expression values were obtained using 'Model-based expression' method and '5th percentile of region (PM-only)' as background using dChip. The genes were filtered according to the variation across samples: 0.20 < Standard deviation/mean < 100.00. 8,835 of 38,439 probe sets satisfied the filtering criteria. In addition, the upper and lower threshold ratio was set to 1.8 while performing pair-wise comparisons. Hierarchical clustering of the gene expression profile was performed using Pearson's correlation with a threshold of $p < 0.005$ for calling significant clusters. Cluster dendrogram was generated in 'R' (www.r-project.org). Three different sets of mouse B-cell data [147-149] were converted into chicken orthologs using the Netaffx®. Data are deposited in GEO with an accession number (GSE12466).

In **Paper II**, raw data were pre-processed using R statistical programming language, Bioconductor (http://www.bioconductor.org/), and the 'affy' package. Data were initially RMA normalized first across the samples and then within each sample group. RMA normalized data were then scaled to a common median value. Both raw and pre-processed data is deposited in GEO (GSE30627). Further filtering, within sample group analyses and pair-wise comparisons were carried out using dChip

(https://sites.google.com/site/dchipsoft/, http://biosun1.harvard.edu/complab/dchip/) and Microsoft Excel.

In **Paper III**, The pre-processing of the data was performed using MAS5.0 algorithm in the Affymetrix Expression Array Console. The pre-processed data were filtered upon the assessed as background value criteria that were set as 50. Assessment of background value was based upon the number of 'Absent call' in all the probe sets. Additional filtering was done after taking away the Affymetrix control gene –AFFX. In total 19,112 probe-sets fulfilled the criteria. Further analyses were carried out using Subio Platform (http://www.subio.jp/). Pairwise comparisons were carried out using different statistical methods in three T cell subsets. For the $CD3^+$ T cells, F-test was performed in order to verify the variance in the replicates, T test of equal variance was chosen as the statistical method with a fold change of $^+$/-2 and false discovery rate (FDR) 0.05. For the $CD4^+$ populations, variance testing methods were applied with a with a fold change of $^+$/-2. In the $CD8^+$ population, since there is only a single sample, only fold change of $^+$/-2 was chosen for differentially expressed genes. In total 10 arrays were analyzed which are accessible through the Gene Expression Omnibus (GEO; GSE12466).

3.7 PCR and DNA sequencing

Conventional PCR technique is a widely used method in molecular biology used for amplifying DNA sequences [150]. PCR products can be directly sequenced by various of methods using fluorophore-labeled sequencing primers [151]. In **Paper II**, four kb (2L:8274950,8279050) surrounding the Btk29A-locus in drosophila genome was used as a template to construct 8 forward-, and 8 reverse oriented primers covering

the entire 4-kb region from both ends with a 500 bp spacing. Primers were also made for the 3´- and 5´-ends of the Btk29A^{ficP} -responsible P-element (BmΔ-w). Both primers were facing outwards from the P-element. PCR was performed using ABI GeneAmpTM system 2700 and the insertion site was determined by sequencing the PCR product (http://www.eurofinsdna.com).

4 RESULTS AND DISCUSSION

4.1 Paper I

Expression profiling of chicken DT40 lymphoma cells indicates clonal selection of knockout and gene reconstituted cells

The DT40 cell-line has been extensively used to create deletion mutants, one of which lacks Bruton's tyrosine kinase (Btk). It was previously shown that there are differences in gene expression between wild-type and Btk-deficient animals [147, 148]. Global gene expression profiling of the avian B-lymphoma DT40 cell-line was used as a model to differentiate among Btk knockout (KO) and Btk KO cells reconstituted with human Btk. We observed that there was substantial difference in the expression pattern between Wt DT40 and Btk KO B7.10 cells. However, the human Btk-reconstituted B41.13 was more similar towards its parental cell-line (Btk KO B7.10) than towards DT40 cells. The same effect was observed in Btk KO cells reconstituted with kinase-inactive human Btk. It is likely that this reflects the much higher degree of clonal selection, which takes place during the generation of deletion mutants as compared to cell-lines reconstituted by transfection. Differences in the gene expression pattern showed statistically significant changes between parental DT40 and all the Btk KO cell populations irrespective of whether they are reconstituted or not. These results imply that in the process of generating a knockout cell-line, sub-clones are selected, which have multiple changes in their gene expression pattern ($p<0.01$). Although other parameters could also influence the expression profile, this potentially has important implications when interpreting microarray data from gene-deleted cell-lines.

DT40 cell-line has been used as a model for the study of proximal signaling events [152, 153]. It has been previously shown that chicken KO cells reconstituted with a

human *BTK* gene have the same functional activity as parental DT40 cells, when analyzing membrane-proximal signaling events [154, 155]. In this study, the cell-lines were compared to each other in order to verify the degree of clonal selection among the deletion mutants. We observed that the clonal selection is the dominant underlying factor, in contrast to proximal events, when transcriptional patterns were investigated using microarray gene expression profiling.

4.2 Paper II

Differential evolutionary wiring of the tyrosine kinase Btk.

Btk29A is considered to be the fly ortholog of Bruton's tyrosine kinase (Btk), a non-receptor-type tyrosine kinase of the Tec family. In contrast to the mammalian Tec gene family that is composed of five members, Drosophila has only one Tec family gene, *Btk29A*, with two splice variants-type 1 and type 2. In fact, the protein product from the type 2 splice variant of this gene exhibits the highest homology to Btk among the five mammalian Tec kinases. The type 1 splice variant has a shorter N-terminus that is unique to *Drosophila* Btk29A. *Btk29A* displays a dynamic pattern of expression through the embryonic to adult stages. $Btk29A^{ficP}$ is a unique allele in that it is devoid of the type 2 isoform while leaving type 1 isoform intact. The $Btk29A^{ficP}$ mutants survive through to the adult stage, exhibiting a copulation defect and reduced lifespan after eclosion. We used Affymetrix® Drosophila Genome 2.0 Genechip arrays containing 18, 770 transcripts - corresponding to the vast majority of known Drosophila genes. We mainly focused on the central nervous tissue as it is known that Btk 29A is mostly expressed in these tissues. It is also known that reduced adult lifespan phenotypes are associated with the selective loss of the type 2 isoform, which represents the Btk ortholog. We carried out micro-dissection of the CNS tissue from both adult and larval

stage of $Btk29A^{ficP}$ flies and the revertant strain $fic^{Exc.1-16}$ in order to check whether the loss-of-function phenotype can be rescued at the transcriptomic level.

Out of 7,004-7,979 transcripts expressed in the four sample groups; 5,587 (70-79%) were found in all four tissues and strains. Here, we investigated the role of Btk29A type 2 on a transcriptomic level in larval CNS and adult heads. We used samples either selectively defective in Btk29A type 2 $Btk29A^{ficP}$ or revertant flies with restored Btk29A type 2-function $fic^{Exc.1-16}$. The whole transcriptomic profile for the different sample groups revealed Gene Ontology patterns reflecting lifespan abnormalities in adult head neuronal tissue, but not in larvae. In the Btk29A type 2-deficient strains there was no significant overlap between transcriptomic alterations in adult heads and larvae neuronal tissue, respectively. Moreover, there was no significant overlap of the transcriptomic changes between flies and mammals, suggesting that the evolutionary conservation is confined to components of the proximal signaling, whereas the corresponding, downstream transcriptional regulation has been differentially wired.

4.3 Paper III

Estimating the influence of CD4+ CD8+ T cell interactions on the transcriptome

T cell activation is one of the important steps in the immune response where hundreds of genes and proteins play crucial roles in a highly organized manner. Activation of those genes in the immune system is dependent on the T cell receptor signaling pathway and its regulation of transcription [156]. Most of the research laboratories follow the more conventional method and separate lymphocytes before activation. In this study, we show that this may not be the optimal way since the separated cells show some altered gene expression profile.

We used mouse Affymetrix® MOE 430_2.0 array containing over 45,000 probe-sets and selected around 10,000 probe-sets those were differentially expressed in compare to un-stimulated cells upon activation in $CD4^+$, $CD8^+$ and $CD3^+$ T cells. Expression of 6% of all genes was found to be dependent on $CD4^+$-$CD8^+$ interactions. Gene set enrichment analysis (GSEA) of these genes revealed 22 overrepresented Gene Ontology (GO) Biological processes. Among those 'adaptive immune response' and 'lymphocyte/leucocyte mediated immunity' were of interest in the context of T cell activation. We also observed similar expression pattern in genes involved in immune response and T cell activation compared to the global expression profiling. . Our results shows that there is has no such major influence on the induced transcriptomes in the context *in vitro* T cell culture when it is separated first and then activated.

5 CONCLUDING REMARKS

Microarray gene expression profiling has been regarded as one the most powerful tools in molecular medicine for more than a decade. A huge amount of data can be generated and analyzed from a single experiment and that can be compared with archived data all across the globe. Our group has been working with expression profiling of Btk deficiency in primary B cells and Itk deficiency in primary T cells which is eventually complemented with this thesis work. Exploring the chicken cell lines and the Drosophila as a model organism has changed our concept of thinking when it comes to transcriptomic activation. Cross-species studies revealed much dissimilarity in different scenarios in the context of Btk deficiency. Regulation of T cell activation has also been addressed in this work, which might help the scientist to design their experiments in future. Technology advances fast with time and sooner or later, easy availability of the RNA sequencing might take over growing field of microarray expression profiling. Works done in this thesis will hopefully add more knowledge for the scientists working with cell lines and cell activation in the context of transcriptional influence.

6 ACKNOWLEDGEMENTS

I am grateful to all of my friends, colleagues and family who have contributed and supported me during all these years and made this dream come true.

I would like to start with ***Professor Edvard Smith***, a man of dignity and personality who gave me the opportunity to come here in Sweden and work in his lab. It was never possible without your tremendous support, enthusiasm making me who I am now. It was something more than a supervisor role that you guided me all through these years developing not only my scientific skills but also a good human being.

Then I would like to thank my main supervisor, ***Jessica Lindvall***, my 'guru' of Bioinformatics. Without your proper guidance and supervision, I would never have been explored so much the vast world of bioinformatics. You made me learn to think critically and developing my expertise in this field. I am indebted to you for all your efforts.

Per Kylsten, the 'fly' guy with enormous energy and interest in cricket. You introduced me to these little creatures those I dissected under your supervision. A lot of basic fly genetics and molecular biology techniques I have learnt from you.

My mentor ***Atiqul Islam***, the person with whom I shared all my problems and you inspired me at every moment I was down. Thanks Atiq bhai, without this journey was never possible to finish.

I would like to express my sincere gratitude to ***Professor Andrej Weintraub***, Studirektor of LABMED for your help and support throughout the years. I would also like to thank ***Marita Ward*** at the administration of LABMED.

All the past and present members of MCG group, especially,

Emelie Blomberg, thanks for introducing me to the lab environment for the very first time and teaching me all the basics of cell culture, RNA isolation and RT-PCR. ***Lotta Asplund,*** thanks for teaching me DNA preparation, PCR and sequencing. ***Abdalla Jama***, a man of never ending enthusiasm, a great teacher. I learned all kind of blots except the 'Southern' blot! from this extra-ordinary person. ***Beston Nore***, thanks for your help in the early days in the lab. ***Leonardo Vargas***, thanks for great scientific discussions while working in your project. ***Maroof Hasan***, my fellow countryman, for your guidance and support***.*** ***Pedro Moreno*** and ***Iulian Opera***, I will miss your company and playing football together in the early golden days.

Rani Faryal, a person with motherly affection, I still miss your delicious cooking. ***Alamdar Hussain***, a true friend, for your utmost help during my labworks. I wish you and your family all the very best in the future. The always smiling ***Dara Mohammad***, you brought new essence in the group. ***Maria Cardona***, thanks for your great hospitality. ***Karin Lundin***, I am really grateful for all your help and supports. ***Anna Berglöf,*** for your expertise and your patience while working in the project together. I would also like to thank ***Oscar Simonson***, ***Sofia Stenler***, ***Samir El-Andaloussi***, ***Eman Jaghloul***, ***Joana Viola***, ***Liang Liu***, ***Manuela Gustafsson, Burcu Bestas, Sylvain Geny, Vladimir Pabon, Abdulrahman Hamasy, Olof Gissberg*** for all these years in the lab!

The administration at KFC, ***Hanna Gador***, I am really thankful for all your supports during these years and ***Merja Häkkinen*** for always being helpful at. All the people at BEA for microarrays, especially, ***Marika Rönholm*** and ***David Brodin*** for your expert support. ***Monica*** and ***Lalla*** in the hematology group for your unconditional love and support. I would like to thank ***Stina Salmi*** at studenthälsan for your inspirational and motivational words.

I am really thankful to my past chiefs ***Catherine Udén***, ***Helena Josefsson*** and ***Magnus Johansson*** for giving me opportunity to work in the clinic.

I would like to take the opportunity to show my gratitude to ***Dora Khala*** and ***Tutul mama*** for your support and encouragements throughout the years. I would also like to thank ***Bachchu mama*** and his family for their initial support in my early days in Sweden. ***Babu mama*** and ***Champa mami*** for all your love and support, bästa kusin '***Musse***'. Special thanks to ***Anwar khalu*** and his family to stand beside me and keep faith on me in my bad days. Special thanks to The 'Bangla' community in Flemingsberg, **Maksud Bhai**, **Adil Bhai, Iqbal bhai, Parvez bhai, Tanveer, Risul Dipu**, **Raisul**, **Shaon** and others. Bangladeshi doctors residing in Sweden namely, **Akhter bhai**, **Monon bhai**, **Jewel Bhai**, **Niaz Bhai**, **Bulu Bhai** and others.

My inspiration my mom, '**Mamoni**' and my dad '**Papa**', I love you guys so much. What I have achieved till now is only for you and I dedicate all my success to you. My loveliest sister **Anita**, thank you 'Apu' for all your supports and I know that one day I will feel proud of you as well. Last but not the least, my 'Sayang' and my son **Aryan Nawaz**, we are the family, and without you my life would have been incomplete.

7 REFERENCES

1. Li H: **Systems genetics in "-omics" era: current and future development**. *Theory in biosciences = Theorie in den Biowissenschaften* 2013, **132**(1):1-16.
2. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J: **A concise guide to cDNA microarray analysis**. *BioTechniques* 2000, **29**(3):548-550, 552-544, 556 passim.
3. Smith CI, Islam TC, Mattsson PT, Mohamed AJ, Nore BF, Vihinen M: **The Tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, Tec, Txk and homologs in other species**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2001, **23**(5):436-446.
4. Ortutay C, Nore BF, Vihinen M, Smith CI: **Phylogeny of Tec family kinases identification of a premetazoan origin of Btk, Bmx, Itk, Tec, Txk, and the Btk regulator SH3BP5**. *Advances in genetics* 2008, **64**:51-80.
5. Gomez-Rodriguez J, Readinger JA, Viorritto IC, Mueller KL, Houghtling RA, Schwartzberg PL: **Tec kinases, actin, and cell adhesion**. *Immunological reviews* 2007, **218**:45-64.
6. Berg LJ, Finkelstein LD, Lucas JA, Schwartzberg PL: **Tec family kinases in T lymphocyte development and function**. *Annu Rev Immunol* 2005, **23**:549-600.
7. Ekman N, Arighi E, Rajantie I, Saharinen P, Ristimaki A, Silvennoinen O, Alitalo K: **The Bmx tyrosine kinase is activated by IL-3 and G-CSF in a PI-3K dependent manner**. *Oncogene* 2000, **19**(36):4151-4158.
8. Hu Q, Davidson D, Schwartzberg PL, Macchiarini F, Lenardo MJ, Bluestone JA, Matis LA: **Identification of Rlk, a novel protein tyrosine kinase with predominant expression in the T cell lineage**. *The Journal of biological chemistry* 1995, **270**(4):1928-1934.
9. Smith CI, Baskin B, Humire-Greiff P, Zhou JN, Olsson PG, Maniar HS, Kjellen P, Lambris JD, Christensson B, Hammarstrom L *et al*: **Expression of Bruton's agammaglobulinemia tyrosine kinase gene, BTK, is selectively down-regulated in T lymphocytes and plasma cells**. *J Immunol* 1994, **152**(2):557-565.
10. Schmidt U, Boucheron N, Unger B, Ellmeier W: **The role of Tec family kinases in myeloid cells**. *International archives of allergy and immunology* 2004, **134**(1):65-78.
11. Yao L, Janmey P, Frigeri LG, Han W, Fujita J, Kawakami Y, Apgar JR, Kawakami T: **Pleckstrin homology domains interact with filamentous actin**. *The Journal of biological chemistry* 1999, **274**(28):19752-19761.
12. Fowell DJ, Shinkai K, Liao XC, Beebe AM, Coffman RL, Littman DR, Locksley RM: **Impaired NFATc translocation and failure of Th2 development in Itk-deficient CD4+ T cells**. *Immunity* 1999, **11**(4):399-409.
13. Tsai YT, Su YH, Fang SS, Huang TN, Qiu Y, Jou YS, Shih HM, Kung HJ, Chen RH: **Etk, a Btk family tyrosine kinase, mediates cellular transformation by linking Src to STAT3 activation**. *Molecular and cellular biology* 2000, **20**(6):2043-2054.
14. Uckun FM: **Bruton's tyrosine kinase (BTK) as a dual-function regulator of apoptosis**. *Biochemical pharmacology* 1998, **56**(6):683-691.
15. Takesono A, Finkelstein LD, Schwartzberg PL: **Beyond calcium: new signaling pathways for Tec family kinases**. *Journal of cell science* 2002, **115**(Pt 15):3039-3048.
16. Yu L, Simonson OE, Mohamed AJ, Smith CI: **NF-kappaB regulates the transcription of protein tyrosine kinase Tec**. *The FEBS journal* 2009, **276**(22):6714-6724.
17. Maruyama T, Nara K, Yoshikawa H, Suzuki N: **Txk, a member of the non-receptor tyrosine kinase of the Tec family, forms a complex with poly(ADP-ribose) polymerase 1 and elongation factor 1alpha and regulates interferon-gamma gene transcription in Th1 cells**. *Clinical and experimental immunology* 2007, **147**(1):164-175.

18. Sacristan C, Schattgen SA, Berg LJ, Bunnell SC, Roy AL, Rosenstein Y: **Characterization of a novel interaction between transcription factor TFII-I and the inducible tyrosine kinase in T cells**. *European journal of immunology* 2009, **39**(9):2584-2595.
19. Kawakami Y, Kitaura J, Hata D, Yao L, Kawakami T: **Functions of Bruton's tyrosine kinase in mast and B cells**. *Journal of leukocyte biology* 1999, **65**(3):286-290.
20. Bao Y, Zheng J, Han C, Jin J, Han H, Liu Y, Lau YL, Tu W, Cao X: **Tyrosine kinase Btk is required for NK cell activation**. *The Journal of biological chemistry* 2012, **287**(28):23769-23778.
21. Shinohara M, Koga T, Okamoto K, Sakaguchi S, Arai K, Yasuda H, Takai T, Kodama T, Morio T, Geha RS *et al*: **Tyrosine kinases Btk and Tec regulate osteoclast differentiation by linking RANK and ITAM signals**. *Cell* 2008, **132**(5):794-806.
22. Islam TC, Smith CI: **The cellular phenotype conditions Btk for cell survival or apoptosis signaling**. *Immunol Rev* 2000, **178**:49-63.
23. Hamada N, Backesjo CM, Smith CI, Yamamoto D: **Functional replacement of Drosophila Btk29A with human Btk in male genital development and survival**. *FEBS letters* 2005, **579**(19):4131-4137.
24. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S *et al*: **Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia**. *Cell* 1993, **72**(2):279-290.
25. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flinter F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M *et al*: **The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases**. *Nature* 1993, **361**(6409):226-233.
26. Rawlings DJ, Saffran DC, Tsukada S, Largaespada DA, Grimaldi JC, Cohen L, Mohr RN, Bazan JF, Howard M, Copeland NG *et al*: **Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice**. *Science* 1993, **261**(5119):358-361.
27. Campana D, Farrant J, Inamdar N, Webster AD, Janossy G: **Phenotypic features and proliferative activity of B cell progenitors in X-linked agammaglobulinemia**. *J Immunol* 1990, **145**(6):1675-1680.
28. Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, de Groot R, van Dongen JJ: **Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children**. *Pediatr Res* 2002, **51**(2):159-168.
29. Fruman DA, Satterthwaite AB, Witte ON: **Xid-like phenotypes: a B cell signalosome takes shape**. *Immunity* 2000, **13**(1):1-3.
30. Nundel K, Busto P, Debatis M, Marshak-Rothstein A: **The role of Bruton's tyrosine kinase in the development and BCR/TLR-dependent activation of AM14 rheumatoid factor B cells**. *Journal of leukocyte biology* 2013.
31. Lindvall JM, Blomberg KE, Valiaho J, Vargas L, Heinonen JE, Berglof A, Mohamed AJ, Nore BF, Vihinen M, Smith CI: **Bruton's tyrosine kinase: cell biology, sequence conservation, mutation spectrum, siRNA modifications, and expression profiling**. *Immunological reviews* 2005, **203**:200-215.
32. Muller S, Maas A, Islam TC, Sideras P, Suske G, Philipsen S, Xanthopoulos KG, Hendriks RW, Smith CI: **Synergistic activation of the human Btk promoter by transcription factors Sp1/3 and PU.1**. *Biochem Biophys Res Commun* 1999, **259**(2):364-369.
33. Webb CF, Yamashita Y, Ayers N, Evetts S, Paulin Y, Conley ME, Smith EA: **The transcription factor Bright associates with Bruton's tyrosine kinase, the defective protein in immunodeficiency disease**. *J Immunol* 2000, **165**(12):6956-6965.
34. Rajaiya J, Hatfield M, Nixon JC, Rawlings DJ, Webb CF: **Bruton's tyrosine kinase regulates immunoglobulin promoter activation in association with**

the transcription factor Bright. *Molecular and cellular biology* 2005, **25**(6):2073-2084.
35. Khan WN: **Regulation of B lymphocyte development and activation by Bruton's tyrosine kinase**. *Immunologic research* 2001, **23**(2-3):147-156.
36. Yang W, Desiderio S: **BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(2):604-609.
37. Schmidt C, Kim D, Ippolito GC, Naqvi HR, Probst L, Mathur S, Rosas-Acosta G, Wilson VG, Oldham AL, Poenie M *et al*: **Signalling of the BCR is regulated by a lipid rafts-localised transcription factor, Bright**. *The EMBO journal* 2009, **28**(6):711-724.
38. Rajaiya J, Nixon JC, Ayers N, Desgranges ZP, Roy AL, Webb CF: **Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I**. *Molecular and cellular biology* 2006, **26**(12):4758-4768.
39. Hirano M, Kikuchi Y, Nisitani S, Yamaguchi A, Satoh A, Ito T, Iba H, Takatsu K: **Bruton's tyrosine kinase (Btk) enhances transcriptional co-activation activity of BAM11, a Btk-associated molecule of a subunit of SWI/SNF complexes**. *Int Immunol* 2004, **16**(5):747-757.
40. Mahajan S, Vassilev A, Sun N, Ozer Z, Mao C, Uckun FM: **Transcription factor STAT5A is a substrate of Bruton's tyrosine kinase in B cells**. *The Journal of biological chemistry* 2001, **276**(33):31216-31228.
41. Baba K, Takeshita A, Majima K, Ueda R, Kondo S, Juni N, Yamamoto D: **The Drosophila Bruton's tyrosine kinase (Btk) homolog is required for adult survival and male genital formation**. *Molecular and cellular biology* 1999, **19**(6):4405-4413.
42. Roulier EM, Panzer S, Beckendorf SK: **The Tec29 tyrosine kinase is required during Drosophila embryogenesis and interacts with Src64 in ring canal development**. *Molecular cell* 1998, **1**(6):819-829.
43. Dodson GS, Guarnieri DJ, Simon MA: **Src64 is required for ovarian ring canal morphogenesis during Drosophila oogenesis**. *Development* 1998, **125**(15):2883-2892.
44. Lu N, Guarnieri DJ, Simon MA: **Localization of Tec29 to ring canals is mediated by Src64 and PtdIns(3,4,5)P3-dependent mechanisms**. *The EMBO journal* 2004, **23**(5):1089-1100.
45. Yamadori T, Baba Y, Matsushita M, Hashimoto S, Kurosaki M, Kurosaki T, Kishimoto T, Tsukada S: **Bruton's tyrosine kinase activity is negatively regulated by Sab, the Btk-SH3 domain-binding protein**. *Proc Natl Acad Sci U S A* 1999, **96**(11):6341-6346.
46. Shinohara H, Kurosaki T: **Genetic analysis of B cell signaling**. *Sub-cellular biochemistry* 2006, **40**:145-187.
47. Tsubata T, Wienands J: **B cell signaling. Introduction**. *Int Rev Immunol* 2001, **20**(6):675-678.
48. Kurosaki T, Hikida M: **Tyrosine kinases and their substrates in B lymphocytes**. *Immunological reviews* 2009, **228**(1):132-148.
49. Kurosaki T: **Regulation of B-cell signal transduction by adaptor proteins**. *Nat Rev Immunol* 2002, **2**(5):354-363.
50. Rawlings DJ, Scharenberg AM, Park H, Wahl MI, Lin S, Kato RM, Fluckiger AC, Witte ON, Kinet JP: **Activation of BTK by a phosphorylation mechanism initiated by SRC family kinases**. *Science* 1996, **271**(5250):822-825.
51. Baba Y, Kurosaki T: **Impact of Ca2+ signaling on B cell function**. *Trends Immunol* 2011, **32**(12):589-594.
52. Liao XC, Littman DR: **Altered T cell receptor signaling and disrupted T cell development in mice lacking Itk**. *Immunity* 1995, **3**(6):757-769.
53. Hogan PG, Chen L, Nardone J, Rao A: **Transcriptional regulation by calcium, calcineurin, and NFAT**. *Genes Dev* 2003, **17**(18):2205-2232.

54. Palacios EH, Weiss A: **Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation**. *Oncogene* 2004, **23**(48):7990-8000.
55. Parsons SJ, Parsons JT: **Src family kinases, key regulators of signal transduction**. *Oncogene* 2004, **23**(48):7906-7909.
56. Salmond RJ, Filby A, Qureshi I, Caserta S, Zamoyska R: **T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance**. *Immunological reviews* 2009, **228**(1):9-22.
57. Seddon B, Zamoyska R: **TCR signals mediated by Src family kinases are essential for the survival of naive T cells**. *J Immunol* 2002, **169**(6):2997-3005.
58. Stefanova I, Dorfman JR, Germain RN: **Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes**. *Nature* 2002, **420**(6914):429-434.
59. Veillette A, Bookman MA, Horak EM, Bolen JB: **The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck**. *Cell* 1988, **55**(2):301-308.
60. Artyomov MN, Lis M, Devadas S, Davis MM, Chakraborty AK: **CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(39):16916-16921.
61. Lovatt M, Filby A, Parravicini V, Werlen G, Palmer E, Zamoyska R: **Lck regulates the threshold of activation in primary T cells, while both Lck and Fyn contribute to the magnitude of the extracellular signal-related kinase response**. *Molecular and cellular biology* 2006, **26**(22):8655-8665.
62. Acuto O, Di Bartolo V, Michel F: **Tailoring T-cell receptor signals by proximal negative feedback mechanisms**. *Nat Rev Immunol* 2008, **8**(9):699-712.
63. Deindl S, Kadlecek TA, Brdicka T, Cao X, Weiss A, Kuriyan J: **Structural basis for the inhibition of tyrosine kinase activity of ZAP-70**. *Cell* 2007, **129**(4):735-746.
64. August A, Gibson S, Kawakami Y, Kawakami T, Mills GB, Dupont B: **CD28 is associated with and induces the immediate tyrosine phosphorylation and activation of the Tec family kinase ITK/EMT in the human Jurkat leukemic T-cell line**. *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**(20):9347-9351.
65. King PD, Sadra A, Teng JM, Xiao-Rong L, Han A, Selvakumar A, August A, Dupont B: **Analysis of CD28 cytoplasmic tail tyrosine residues as regulators and substrates for the protein tyrosine kinases, EMT and LCK**. *J Immunol* 1997, **158**(2):580-590.
66. Chambers CA: **The expanding world of co-stimulation: the two-signal model revisited**. *Trends Immunol* 2001, **22**(4):217-223.
67. Parry RV, Riley JL, Ward SG: **Signalling to suit function: tailoring phosphoinositide 3-kinase during T-cell activation**. *Trends Immunol* 2007, **28**(4):161-168.
68. Wallner S, Gruber T, Baier G, Wolf D: **Releasing the brake: targeting Cbl-b to enhance lymphocyte effector functions**. *Clinical & developmental immunology* 2012, **2012**:692639.
69. Bernard A, Lamy, Alberti I: **The two-signal model of T-cell activation after 30 years**. *Transplantation* 2002, **73**(1 Suppl):S31-35.
70. Readinger JA, Mueller KL, Venegas AM, Horai R, Schwartzberg PL: **Tec kinases regulate T-lymphocyte development and function: new insights into the roles of Itk and Rlk/Txk**. *Immunological reviews* 2009, **228**(1):93-114.
71. Weiss A, Imboden JB: **Cell surface molecules and early events involved in human T lymphocyte activation**. *Adv Immunol* 1987, **41**:1-38.
72. Bockenstedt LK, Goldsmith MA, Koretzky GA, Weiss A: **The activation of T lymphocytes**. *Rheum Dis Clin North Am* 1987, **13**(3):411-430.

73. Havran WL, Poenie M, Kimura J, Tsien R, Weiss A, Allison JP: **Expression and function of the CD3-antigen receptor on murine CD4+8+ thymocytes**. *Nature* 1987, **330**(6144):170-173.
74. Burmester GR, Jahn B, Gramatzki M, Zacher J, Kalden JR: **Activated T cells in vivo and in vitro: divergence in expression of Tac and Ia antigens in the nonblastoid small T cells of inflammation and normal T cells activated in vitro**. *J Immunol* 1984, **133**(3):1230-1234.
75. Dietrich PY, Walker PR, Schnuriger V, Saas P, Perrin G, Guillard M, Gaudin C, Caignard A: **TCR analysis reveals significant repertoire selection during in vitro lymphocyte culture**. *Int Immunol* 1997, **9**(8):1073-1083.
76. Tanaka H, Taniuchi I: **The CD4/CD8 Lineages: Central Decisions and Peripheral Modifications for T Lymphocytes**. *Current topics in microbiology and immunology* 2013.
77. Naito T, Taniuchi I: **The network of transcription factors that underlie the CD4 versus CD8 lineage decision**. *International immunology* 2010, **22**(10):791-796.
78. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic Acids Res* 2002, **30**(1):207-210.
79. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nature reviews Genetics* 2009, **10**(1):57-63.
80. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T *et al*: **Direct multiplexed measurement of gene expression with color-coded probe pairs**. *Nature biotechnology* 2008, **26**(3):317-325.
81. Heller MJ: **DNA microarray technology: devices, systems, and applications**. *Annu Rev Biomed Eng* 2002, **4**:129-153.
82. Ozsolak F, Milos PM: **RNA sequencing: advances, challenges and opportunities**. *Nature reviews Genetics* 2011, **12**(2):87-98.
83. Martin JA, Wang Z: **Next-generation transcriptome assembly**. *Nature reviews Genetics* 2011, **12**(10):671-682.
84. Pareek CS, Smoczynski R, Tretyn A: **Sequencing technologies and genome sequencing**. *Journal of applied genetics* 2011, **52**(4):413-435.
85. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L: **Next-generation sequencing and its applications in molecular diagnostics**. *Expert review of molecular diagnostics* 2011, **11**(3):333-343.
86. Shiroguchi K, Jia TZ, Sims PA, Xie XS: **Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(4):1347-1352.
87. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hossein A, Fare T: **Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System**. *BMC research notes* 2009, **2**:80.
88. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**(5235):467-470.
89. Shalon D, Smith SJ, Brown PO: **A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization**. *Genome research* 1996, **6**(7):639-645.
90. Xiang CC, Chen Y: **cDNA microarray technology and its applications**. *Biotechnology advances* 2000, **18**(1):35-46.
91. Haab BB: **Methods and applications of antibody microarrays in cancer research**. *Proteomics* 2003, **3**(11):2116-2122.
92. Angenendt P: **Progress in protein and antibody microarray technology**. *Drug Discov Today* 2005, **10**(7):503-511.
93. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G: **Making and reading microarrays**. *Nature genetics* 1999, **21**(1 Suppl):15-19.

94. Petersen DW, Kawasaki ES: **Manufacturing of microarrays**. *Adv Exp Med Biol* 2007, **593**:1-11.
95. Urakawa H, Noble PA, El Fantroussi S, Kelly JJ, Stahl DA: **Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses**. *Applied and environmental microbiology* 2002, **68**(1):235-244.
96. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays**. *Nature genetics* 1999, **21**(1 Suppl):20-24.
97. McGall GH, Christians FC: **High-density genechip oligonucleotide probe arrays**. *Advances in biochemical engineering/biotechnology* 2002, **77**:21-42.
98. Wolber PK, Collins PJ, Lucas AB, De Witte A, Shannon KW: **The Agilent in situ-synthesized microarray platform**. *Methods in enzymology* 2006, **410**:28-57.
99. Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, Lebruska LL, Laurent M, Shen R, Barker D: **Illumina universal bead arrays**. *Methods in enzymology* 2006, **410**:57-73.
100. Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J *et al*: **Gene expression analysis using oligonucleotide arrays produced by maskless photolithography**. *Genome research* 2002, **12**(11):1749-1755.
101. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H *et al*: **Expression monitoring by hybridization to high-density oligonucleotide arrays**. *Nature biotechnology* 1996, **14**(13):1675-1680.
102. .
103. Schinke-Braun M, Couget JA: **Expression profiling using affymetrix genechip probe arrays**. *Methods Mol Biol* 2007, **366**:13-40.
104. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG: **The affymetrix GeneChip platform: an overview**. *Methods in enzymology* 2006, **410**:3-28.
105. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH: **Amplified RNA synthesized from limited quantities of heterogeneous cDNA**. *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(5):1663-1667.
106. Arteaga-Salas JM, Zuzan H, Langdon WB, Upton GJ, Harrison AP: **An overview of image-processing methods for Affymetrix GeneChips**. *Briefings in bioinformatics* 2008, **9**(1):25-33.
107. Jiang N, Leach LJ, Hu X, Potokina E, Jia T, Druka A, Waugh R, Kearsey MJ, Luo ZW: **Methods for evaluating gene expression from Affymetrix microarray datasets**. *BMC bioinformatics* 2008, **9**:284.
108. Ahmed FE: **Microarray RNA transcriptional profiling: part II. Analytical considerations and annotation**. *Expert review of molecular diagnostics* 2006, **6**(5):703-715.
109. Qu Y, He F, Chen Y: **Different effects of the probe summarization algorithms PLIER and RMA on high-level analysis of Affymetrix exon arrays**. *BMC bioinformatics* 2010, **11**:211.
110. Schadt EE, Edwards SW, GuhaThakurta D, Holder D, Ying L, Svetnik V, Leonardson A, Hart KW, Russell A, Li G *et al*: **A comprehensive transcript index of the human genome generated using microarrays and computational approaches**. *Genome biology* 2004, **5**(10):R73.
111. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data**. *Nucleic Acids Res* 2003, **31**(4):e15.
112. **Affymetrix (2002) Statistical Algorithms Description Document http://www.affymetrix.com/support/technical/whitepapers.pdf.**
113. Li C, Wong WH: **Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(1):31-36.

114. Li C, Hung Wong W: **Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application**. *Genome biology* 2001, **2**(8):RESEARCH0032.
115. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data**. *Biostatistics* 2003, **4**(2):249-264.
116. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**(2):185-193.
117. Wu Z, Irizarry RA: **Preprocessing of oligonucleotide array data**. *Nature biotechnology* 2004, **22**(6):656-658; author reply 658.
118. Quackenbush J: **Computational analysis of microarray data**. *Nature reviews Genetics* 2001, **2**(6):418-427.
119. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression**. *Methods in enzymology* 1999, **303**:179-205.
120. Yeung KY, Ruzzo WL: **Principal component analysis for clustering gene expression data**. *Bioinformatics* 2001, **17**(9):763-774.
121. Raychaudhuri S, Stuart JM, Altman RB: **Principal components analysis to summarize microarray experiments: application to sporulation time series**. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing* 2000:455-466.
122. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(9):5116-5121.
123. Lin D, Shkedy Z, Burzykowski T, Ion R, Gohlmann HW, Bondt AD, Perer T, Geerts T, Van den Wyngaert I, Bijnens L: **An investigation on performance of Significance Analysis of Microarray (SAM) for the comparisons of several treatments with one control in the presence of small-variance genes**. *Biometrical journal Biometrische Zeitschrift* 2008, **50**(5):801-823.
124. Benjamini Y, and Y. Hochberg: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc* 1995, **Ser. B 57 (1): 289–300.**
125. Rice TK, Schork NJ, Rao DC: **Methods for handling multiple testing**. *Advances in genetics* 2008, **60**:293-308.
126. Dalmasso C, Broet P, Moreau T: **[Multiple comparison procedures: principles, limits. Applications to microarray phenotype-genotype analysis]**. *Revue d'epidemiologie et de sante publique* 2004, **52**(6):523-537.
127. Storey JD, Tibshirani R: **Statistical methods for identifying differentially expressed genes in DNA microarrays**. *Methods Mol Biol* 2003, **224**:149-157.
128. Beissbarth T: **Interpreting experimental results using gene ontologies**. *Methods in enzymology* 2006, **411**:340-352.
129. Hung JH, Yang TH, Hu Z, Weng Z, DeLisi C: **Gene set enrichment analysis: performance evaluation and usage guidelines**. *Briefings in bioinformatics* 2012, **13**(3):281-291.
130. Sonoda E, Morrison C, Yamashita YM, Takata M, Takeda S: **Reverse genetic studies of homologous DNA recombination using the chicken B-lymphocyte line, DT40**. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2001, **356**(1405):111-117.
131. Kurosaki T, Maeda A, Ishiai M, Hashimoto A, Inabe K, Takata M: **Regulation of the phospholipase C-gamma2 pathway in B cells**. *Immunological reviews* 2000, **176**:19-29.
132. Shortman K, Cerottini JC, Brunner KT: **The separation of sub-populations of T and B lymphocytes**. *European journal of immunology* 1972, **2**(4):313-319.
133. Shortman K, Byrd WJ, Cerottini JC, Brunner KT: **Characterisation and separation of mouse lymphocyte subpopulations responding to phytohemagglutinin and pokeweed mitogens**. *Cellular immunology* 1973, **6**(1):25-40.

134. Renzi P, Ginns LC: **Analysis of T cell subsets in normal adults. Comparison of whole blood lysis technique to Ficoll-Hypaque separation by flow cytometry**. *Journal of immunological methods* 1987, **98**(1):53-56.
135. Loken MR, Herzenber LA: **Analysis of cell populations with a fluorescence-activated cell sorter**. *Annals of the New York Academy of Sciences* 1975, **254**:163-171.
136. Clark NW, Parkhouse RM, Simmonds RG: **Positive and negative selection of cells by hapten-modified antibodies**. *Journal of immunological methods* 1982, **51**(2):167-170.
137. Blomberg KE, Boucheron N, Lindvall JM, Yu L, Raberger J, Berglof A, Ellmeier W, Smith CE: **Transcriptional signatures of Itk-deficient CD3+, CD4+ and CD8+ T-cells**. *BMC genomics* 2009, **10**:233.
138. Chiao JW, Good RA: **Isolation of human T lymphocytes by a negative selection procedure**. *Immunological communications* 1979, **8**(1):49-54.
139. Rubin GM, Lewis EB: **A brief history of Drosophila's contributions to genome research**. *Science* 2000, **287**(5461):2216-2218.
140. Arias AM: **Drosophila melanogaster and the development of biology in the 20th century**. *Methods Mol Biol* 2008, **420**:1-25.
141. Pandey UB, Nichols CD: **Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery**. *Pharmacological reviews* 2011, **63**(2):411-436.
142. Gonzalez C: **Drosophila melanogaster: a model and a tool to investigate malignancy and identify new therapeutics**. *Nature reviews Cancer* 2013, **13**(3):172-183.
143. Inoue H, Kondo S, Hinohara Y, Juni N, Yamamoto D: **Enhanced phosphorylation and enzymatic activity of phosphoglucomutase by the Btk29A tyrosine kinase in Drosophila**. *Archives of biochemistry and biophysics* 2003, **413**(2):207-212.
144. Desjardins P, Conklin D: **NanoDrop microvolume quantitation of nucleic acids**. *Journal of visualized experiments : JoVE* 2010(45).
145. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements**. *BMC molecular biology* 2006, **7**:3.
146. Cope L, Hartman SM, Gohlmann HW, Tiesman JP, Irizarry RA: **Analysis of Affymetrix GeneChip data using amplified RNA**. *BioTechniques* 2006, **40**(2):165-166, 168, 170.
147. Lindvall JM, Blomberg KE, Berglof A, Yang Q, Smith CI, Islam TC: **Gene expression profile of B cells from Xid mice and Btk knockout mice**. *European journal of immunology* 2004, **34**(7):1981-1991.
148. Lindvall JM, Blomberg KE, Berglof A, Smith CI: **Distinct gene expression signature in Btk-defective T1 B-cells**. *Biochemical and biophysical research communications* 2006, **346**(2):461-469.
149. Hasan M, Lopez-Herrera G, Blomberg KE, Lindvall JM, Berglof A, Smith CI, Vargas L: **Defective Toll-like receptor 9-mediated cytokine production in B cells from Bruton's tyrosine kinase-deficient mice**. *Immunology* 2008, **123**(2):239-249.
150. Bell J: **The polymerase chain reaction**. *Immunology today* 1989, **10**(10):351-355.
151. McBride LJ, Koepf SM, Gibbs RA, Salser W, Mayrand PE, Hunkapiller MW, Kronick MN: **Automated DNA sequencing methods involving polymerase chain reaction**. *Clinical chemistry* 1989, **35**(11):2196-2201.
152. Korade-Mirnics Z, Corey SJ: **Src kinase-mediated signaling in leukocytes**. *Journal of leukocyte biology* 2000, **68**(5):603-613.
153. Perraud AL, Schmitz C, Scharenberg AM: **Calcium signaling, ion channels and more. The DT40 system as a model of vertebrate ion homeostasis and cell physiology**. *Sub-cellular biochemistry* 2006, **40**:257-270.

154. Tomlinson MG, Kurosaki T, Berson AE, Fujii GH, Johnston JA, Bolen JB: **Reconstitution of Btk signaling by the atypical tec family tyrosine kinases Bmx and Txk**. *The Journal of biological chemistry* 1999, **274**(19):13577-13585.
155. Petro JB, Khan WN: **Phospholipase C-gamma 2 couples Bruton's tyrosine kinase to the NF-kappaB signaling pathway in B lymphocytes**. *The Journal of biological chemistry* 2001, **276**(3):1715-1719.
156. Huse M: **The T-cell-receptor signaling network**. *Journal of cell science* 2009, **122**(Pt 9):1269-1273.

Biochemical and Biophysical Research Communications 377 (2008) 584–588

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Expression profiling of chicken DT40 lymphoma cells indicates clonal selection of knockout and gene reconstituted cells

Hossain M. Nawaz [a,*,1], K. Emelie M. Blomberg [a,1], Jessica M. Lindvall [b], Tomohiro Kurosaki [c], C.I. Edvard Smith [a,*]

[a] Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, 14186 Huddinge, Stockholm, Sweden
[b] Department of Informatics, University of Oslo, Oslo, Norway
[c] Laboratory for Lymphocyte Differentiation, RIKEN Research Center for Allergy and Immunology, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

ARTICLE INFO

Article history:
Received 3 October 2008
Available online 16 October 2008

Keywords:
Bruton's tyrosine kinase
B cells
Expression profiling
DT40
X-linked agammaglobulinemia

ABSTRACT

The DT40 cell-line has been extensively used to create deletion mutants, one of which lacks Bruton's tyrosine kinase (Btk). Btk is a cytoplasmic tyrosine kinase important for B-lymphocyte maturation. It was previously shown that there are differences in gene expression between wild-type and Btk-deficient animals. Global gene expression profiling of the avian B-lymphoma DT40 cell-line was used as a model to differentiate among Btk knockout (KO) and Btk KO cells reconstituted with human Btk. Differences in the gene expression pattern showed statistically significant changes between parental DT40 and all the Btk KO cell populations irrespective of whether they are reconstituted or not. These results imply that in the process of generating a knockout cell-line, sub-clones are selected, which have multiple changes in their gene expression pattern ($p < 0.01$). Although other parameters could also influence the expression profile, this potentially has important implications when interpreting microarray data from gene-deleted cell-lines.

Mutations affecting Bruton's tyrosine kinase (Btk) cause disruption in B-lymphocyte development, leading to X-linked agammaglobulinemia (XLA) in humans [1,2] and the less severe X-linked immunodeficiency *(Xid)* in mice [3,4]. Mice deficient in both Btk and the related kinase Tec have an XLA-like phenotype [5]. Btk is a non-receptor tyrosine kinase belonging to the Tec family [6–8]. Btk features, from the N-terminal, a PH domain, a Tec homology (TH) and Src homology domains 3 (SH3), 2 (SH2), and 1 (SH1) [9]. Mutations affecting any of Btk's domains cause XLA and the striking feature of B-cells lacking functional Btk is that they fail to expand. Btk is highly conserved during evolution appearing in the unicellular choanoflagellate *Monosiga brevicollis* [10] and also found in *Drosophila melanogaster* [11] and other vertebrates. A sign of the conservation is the fact that the human *BTK* gene can partly substitute for loss-of-function of the *D. melanogaster Btk29A gene* [12].

Expression profiling has enhanced the understanding of cellular signaling molecules and their downstream targets in health and disease. Btk-deficiency is ideal for expression profiling, since the defect is intrinsic to the B-cell lineage. To determine effects on the transcriptome, we used the Affymetrix Genechip® microarray technology to compare avian DT40 cells with Btk KO as well as deletion mutants reconstituted with Btk. Overt differences in the expression pattern were observed. However, they did not segregate according to the presence and absence of functional Btk, but instead as a result of clonal selection. This potentially has important implications for the use of microarray analysis of cell-lines and suggests that this issue should be subject to further studies.

Materials and methods

Cell-lines: DT40 cell-lines were generated in Dr. T. Kurosaki's laboratory at RIKEN Research Center for Allergy and Immunology in Japan [13]. The cell-lines used in this study were the parental DT40, Btk knockout, B7.10; Btk KO reconstituted with human wild-type Btk, B41.13 or kinase-inactive Btk, B46.5.

Cell culture: The DT40 cells were cultured in RPMI 1640 Glutamax medium with 10% FBS, 1% chicken serum, Penicillin-Streptomycin, 4 mM glutamine, and 50 μM 2-mercaptoethanol (2-ME). DT40 and B7.10 cells were cultivated in three different fetal calf serum batches, while B41.13 and B46.5 were grown in a single serum batch. Both un-stimulated and stimulated cells were grown in the same way, starting in 1 ml serum-free medium for 15 min at 37 °C with 5% CO_2. Stimulations were during that time done with 5 μl supernatant of the hybridoma cell-line M4 (a kind gift from Dr Max D. Cooper, University of Birmingham, AL). After 15 min incubation, 49 ml medium was added and cells were cultivated for 2, 6 or 24 h with, or without, stimuli. In total the experiment was repeated ten times and replicates were pooled prior to expression profiling.

* Corresponding author. Fax: +46 8 58583650.
E-mail addresses: hossain.nawaz@ki.se (H.M. Nawaz), edvard.smith@ki.se (C.E. Smith).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.10.040

RNA isolation and oligonucleotide array hybridization: Total RNA was extracted with RNeasy Mini Kit (Qiagen, Valencia, CA, USA). The RNA amount and the quality were measured with a NanoDrop spectrophotometer (Wilmington, USA) and an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Five micrograms from each of the ten replicates (per sample) were pooled and hybridized to Chicken Genome arrays Genechips®; in total 28 arrays were run. The cRNA synthesis and hybridizations were performed in the BEA core facility at Department of Biosciences, Karolinska Institutet at Novum, Huddinge, Sweden.

Data and statistical analysis strategy for gene selection: In order to optimize the use of microarray chips, all the 28 samples were generated as a pool of 10 different replicates in each [14–16]. Quality control (QC) was checked using Affymetrix® Expression Console software. In order to find distinct patterns in our data set, DNA-Chip Analyzer "dChip" (www.dchip.org) was exercised [17]. dChip uses an intensity-modeling approach for data analysis and clusters microarray probe sets exhibiting similar expression profiles. Data were normalized across all the arrays using probes of 'invariant set' and 'Running median' smoothing. Model-based expression values were obtained using 'Model-based expression' method and '5th percentile of region (PM-only)' as background. The genes were filtered according to the variation across samples: 0.20 < Standard deviation/mean < 100.00. 8,835 of 38,439 probe sets satisfied the filtering criteria. In addition, the upper and lower threshold ratio was set to 1.8 while performing pair-wise comparisons. Hierarchical clustering of the gene expression profile was performed using Pearson's correlation with a threshold of $p < 0.005$ for calling significant clusters. Cluster dendrogram was generated in 'R' (www.r-project.org). Three different sets of mouse B-cell data were converted into chicken orthologs using the Netaffx® (www.affymetrix.com). Ensembl (www.ensembl.org), a public database, was used to find out the number of probe sets used by Affymetrix for each individual gene or transcript for this specific array.

Results

Global gene expression pattern in wild-type DT40 and different Btk mutants

The datasets were initially validated by the presence of internal controls, neomycin phosphotransferase, used for selection as a positive marker for Btk KO cells and Btk itself as a negative marker (Supplementary Fig. 1). The expression of neomycin phosphotransferase was influenced by cell culturing, being reduced 0.3-fold at 24 h compared to 6 h of culture, while the decreased signal from the *Btk* transcripts did not differ significantly (notably, transcripts from the transfected human *Btk* genes are undetectable by avian probe-sets). Unsupervised clustering showed that the B7.10 and the wild-type (WT) B41.13 and kinase-inactive B46.5 reconstituted variants segregated into the same arm, while the parental DT40 formed its own (Fig. 1). Thus, all KO-derived cell-lines, irrespective of reconstitution, showed a more similar expression pattern as compared to the parental DT40 cells. This led us to investigate the clonal effect of the populations used for the array study.

Five hundred and forty two probe-sets were found differentially expressed between DT40 and B7.10. Those specific probe-sets were chosen as selection markers in order to reveal the specific expression pattern in B41.13 and B46.5, respectively. A set of 49 transcripts, consisting of 62 probe-sets, was found to be very similar. Among those, Insulin like growth factor 2 mRNA binding protein 3 (IGF2BP3) was up-regulated in all the deletion mutants. On the other hand, as many as 48 transcripts were found down-regulated at the mRNA expression level (Fig. 2; Supplementary Table I). Thus, the number of transcripts that was down-regulated was much higher. The Affymetrix probe-set ID was used to track the

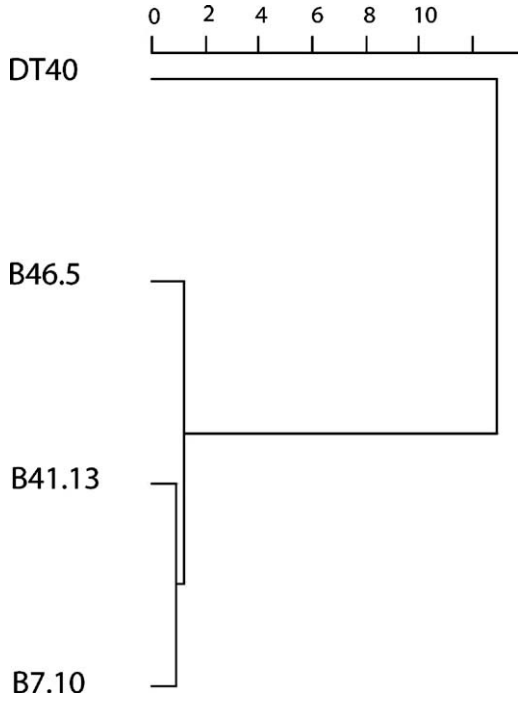

Fig. 1. Clustering pattern of DT40 and Btk KO mutants. Specific pattern of clustering among the mutants.

total number of probe-sets used for those 49 transcripts, representing 49 different genes present in our expression data. Among those, *IGF2BP3* and *USH2A* were found several times. Using the Ensembl database for *Gallus gallus,* we identified 8 transcripts in the correct orientation for *USH2A* and 4 probe-sets for *IGF2BP3* that were included on the array. Among those, 4 probe-sets for *USH2A* and 3 probe-sets for *IGF2BP3* were affected in our data sets. The locations of those specific probe-sets were identified and aligned along the corresponding gene (Fig. 3). One important observation here was that the affected probe-sets behaved in the same way. For example, in *USH2A*, out of the 8 probe-sets, those significantly reduced in Btk KO were all located at one end of the gene (Fig. 3). This distribution was highly significantly different (p value = 0.014) from a random location of the probe-sets.

The cell-lines were compared to each other in order to verify the degree of clonal selection among the deletion mutants. The number of differentially expressed transcripts in each comparison was graphically presented in the form of "inverse" Venn diagrams depicting non-overlapping transcripts (Fig. 4). In DT40 and B7.10 comparisons, the number of differentially expressed transcripts was 903 and 1938 after 6 and 24 h of cultivation, respectively. On the other hand, among the differentially expressed transcripts between Btk KO (B7.10) cells and such cells reconstituted with either WT or kinase-inactive Btk, the number was reduced to a similar extent. In each comparison, including both time points, fewer transcripts differed as to the number in DT40 versus B7.10. Thus, in the 3 comparisons made (panels C and D) an average of 705 and 1158 differentially expressed transcripts were found after 6 and 24 h, respectively.

Cross-validation of the gene expression study

A comparative strategy was taken to cross-validate the hypothesis that clonal changes were the origin of differentially expressed transcripts rather than the presence or absence of Btk. A list of 878 differentially expressed probe-sets was generated with a fold change cut-off value of 2 between DT40 and B7.10 after 6 h of cul-

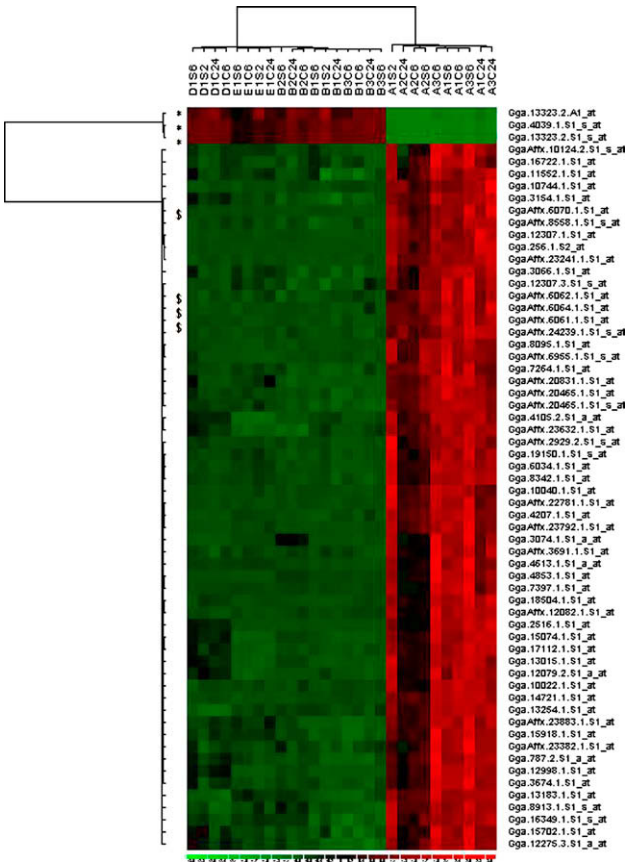

Fig. 2. Expression profiling of DT40 and Btk KO mutants. Heat map showing the expression pattern of the 49 genes, corresponding 62 probe-sets, found significantly changed among all B7.10 and B7.10-reconstituted sub-clones compared to parental DT40. (A = DT40, B = B7.10, D = B41.13, E = B46.5; 1, 2 and 3 are different batches of fetal calf serum; S = Stimulated cells and C = Control, non-stimulated cells. The number at the end corresponds to the number of hours during which the cells were cultivated for (2, 6 and 24 h respectively). *showing *IGF2BP3* and $showing *USH2A* with their corresponding probe-sets.

ture without any stimulation. This list was compared with the differentially expressed genes between DT40 versus B41.13 (referred to as comparison 1) and also with DT40 versus B46.5 (comparison 2). The number of transcripts was more or less similar at different fold changes, ranging from 1.2–2, in both the comparisons mentioned above (Supplementary Table IIA). An *in silico* titration graph was generated and the number of transcripts that changed gradually along with the increase in fold change was similar in both comparisons (Supplementary Fig. 2).

We also compared the anti-IgM stimulated cells with their un-stimulated counterpart to detect any significant difference in between them. Similar comparisons were made in DT40 and B7.10 cell-lines, where a list of differentially expressed genes was generated. This list of transcripts was compared against B41.13 (comparison 3) and B46.5 (comparison 4), respectively. The number of transcripts that overlapped corresponds to the number of transcripts that was found in comparison 1 and 2 (Supplementary Table IIB).

In order to correlate our results to previous studies, data sets were obtained from mouse Btk-defective, transitional type 1 (T1) B-cells [18], and primary splenic B-cells [19], respectively. Both of these studies were conducted using Affymetrix-based microarray chips to identify differentially expressed genes in Btk-deficient B-cells. In both cases, differentially expressed mouse transcripts in between normal and Btk- defective B-cells were converted according to the Affymetrix probe ID into chicken orthologs. Only three probe-sets from the data of Btk-defective T1 B-cells and not a single one from the data of primary splenic B-lymphocytes overlapped with either Comparison 1 or Comparison 2.

Another data set from Btk KO mouse primary spleen B-cells [20] was also included for cross-validation. In this study B-lymphocytes, with or without Btk, and subsequently stimulated with CpG were used for the comparison. A list of differentially expressed transcripts was generated. This list was compared against the chicken data converted into mouse orthologs. In the mouse CpG data, the number of transcripts overlapping between mouse orthologs of chicken data was the same as in the original comparison 1. It was also similar in the case of comparison 2, when mouse orthologs of chicken transcripts between DT40 and B46.5 were compared against the original mouse data. Similar changes were also observed following anti-IgM stimulation of chicken cells (comparison 3 and 4) when transcripts were converted into mouse orthologs and cross-checked against mouse CpG data sets.

We selected these data sets because the cells used in those microarray studies were either in resting state or activated. Comparing our results with data sets from Btk-defective T1 B-cells and primary splenic B-lymphocytes (both at resting state) revealed no similarity or overlap. As cell-lines are considered more like activated cells, higher degree of relatedness was observed in our chicken data when converting the mouse CpG data sets into mouse orthologs and thereafter comparing them to chicken data. Thus, comparison 1 and 2 in chicken cell-lines revealed an average of 62 transcripts that were either up- or down-regulated. A similar value was obtained in the comparison to CpG-stimulated mouse cells (average 64 transcripts). In contrast, as mentioned, when resting cells were used in those comparisons, fewer transcripts were identified *(vide supra).*

Influence of different fetal calf serum batches on expression profiling

Three different batches of fetal calf serum were randomly chosen to study their effects on the cell-lines. It is known that cells show variable growth pattern using different batches of serum, but the extent of variation as measured as expression profiling has to our knowledge not been addressed before. In order to investigate the expression data, we compared the influence of the three different batches (assigned as 1, 2 and 3) of fetal calf serum that were used to supplement the DT40 and B7.10 cell-lines. Cluster dendrogram of the 20 samples that were used for the study of influence of fetal calf serum batches showed that different batches did not cluster together. Instead, they segregated randomly without a definite clustering pattern (Supplementary Fig. 3) with a p value >0.05.

Discussion

Gene expression profiling using microarray has been considered as one of the most valuable tools in molecular biology in recent days. The chicken lymphoma cell-line DT40 has been extensively used owing to its propensity for the generation of deletion mutants (a Medline search October 2, 2008 identified 559 Refs.). These cells have served as a highly useful model for the study of proximal signaling events. It has been previously demonstrated that chicken KO cells reconstituted with a human *BTK* gene have the same functional activity as parental DT40 cells, when analyzing membrane-proximal signaling events [21,22]. The aim was to study the effect of a single-gene mutation in a cell-line and subsequently, the effect of the gene reconstitution. Thus, sub-lines of Btk KO DT40 cells reconstituted with human WT Btk as well as kinase-inactive Btk were studied. However, in contrast to proximal events, we observed that when transcriptional patterns were investigated, clonal selection is suggested to be the dominant underlying factor.

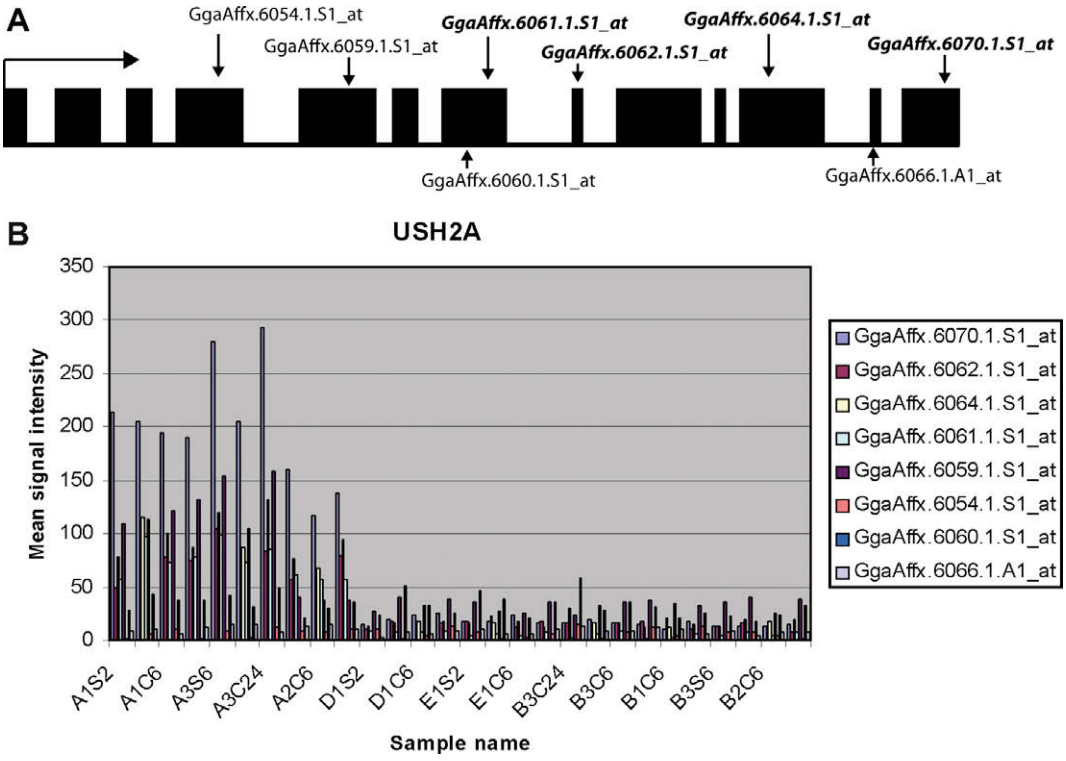

Fig. 3. Location and signals from probe-sets for *USH2A* altered in the expression study. Transcript alignment shows location of the 8 probe-sets for the *USH2A* gene, where the 4 altered transcripts is in italic (panel A). In panel B, bar-plot showing signal intensities of the 8 probe-sets used by Affymetrix for USH2A in the expression data. (A = DT40, B = B7.10, D = B41.13, E = B46.5; 1, 2 and 3 are different batches of fetal calf serum; S=Stimulated cells and C = Control, non-stimulated cells. The number at the end corresponds to the number of hours during which the cells were cultivated for 2, 6 and 24 h, respectively).

In most of the microarray studies of gene defects, a list of differentially expressed genes can be obtained when comparing a WT and a gene KO cell population. However, in cell-lines it is rather difficult to predict whether all of those differentially expressed genes are secondary to the specific gene knockout. A way to verify is to reconstitute the KO population with the deleted gene. We observed that there was substantial difference in the expression pattern between DT40 and B7.10 cells. However, the human Btk-reconstituted B41.13 was more similar towards its parental cell-line (Btk KO B7.10) than towards DT40 cells. The same effect was observed in Btk KO cells reconstituted with kinase-inactive human Btk. It is likely that this reflects the much higher degree of clonal selection, which takes place during the generation of deletion mutants as compared to cell-lines reconstituted by transfection.

Five hundred and forty two transcripts were found to be significantly up-or down-regulated when DT40 was compared to B7.10. 49 genes behaved the same in all three Btk KO mutants, irrespective of whether they were reconstituted or not. Among those,

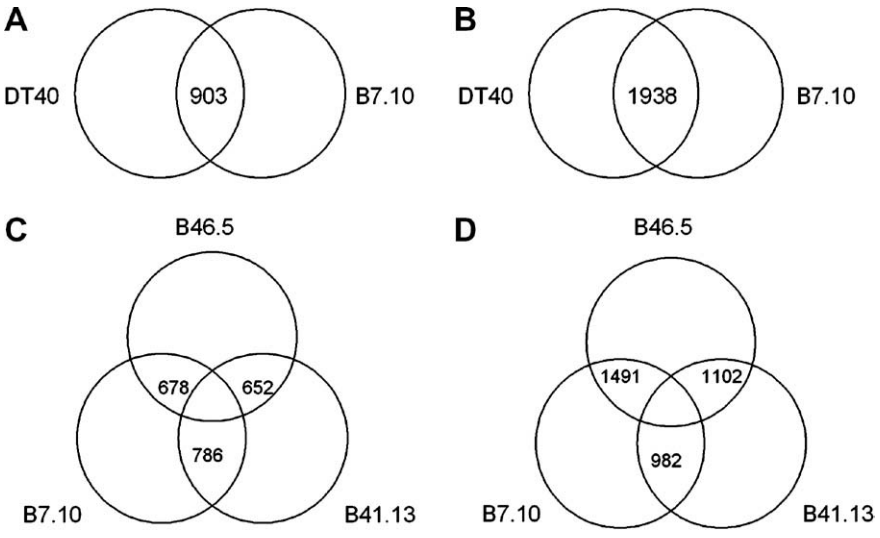

Fig. 4. Differentially expressed transcripts. "Inverse" Venn diagram showing number of differentially expressed transcripts in DT40 and the Btk KO mutants. 903 and 1938 transcripts differ in between DT40 and B7.10 after 6 (panel A) and 24 h (panel B) of cultivation respectively. Panel C and D shows the number of transcripts that differ among the Btk KO mutants in those same time points.

the level of Insulin like growth factor2 mRNA binding protein 3 (*IGF2BP3*) was found up-regulated in all the deletion mutants. A ratio of 48:1 is compatible with the fact that random mutations are expected to mainly induce changes causing loss-of-function. Moreover, using microarray analysis, detectable loss-of-function only reflects reduction at the mRNA level caused by stop-codon formation in the 5′-end of a gene. For these reasons we believe that if loss-of-function effects were measured at the protein level, the alteration would be even more pronounced.

Moreover, an internal validation of our predictions comes from the fact that the same transcripts were identified using independent probe-sets and the further analysis of these. Thus, mRNA from the single gene with increased expression, *IGF2BP3*, was identified thrice, whereas among those with reduced expression, *Usherin syndrome 2A* and *Sorbin and SH3 domain containing 2* were identified by 8 and 3 probe-sets, respectively. For the *USH2A*, the location of target sequences of the corresponding probe-sets showed a non-random distribution. Thus, they formed an uninterrupted probe-cluster. This distribution is compatible with a single deletion encompassing a contiguous stretch of the gene. However, owing to that 3 probe-sets identifying RNA transcribed in the opposite direction were not affected, it is more likely that a non-deletion mutation is the underlying cause. It is interesting to note that there is no mentioning of the possibility of clone selection bias in a previous publication on expression profiling of gene-inactivated DT40 cells [23], while we believe that it is likely to be a general phenomenon in gene-modified cell-lines, owing to the pre-existing clonal variability. Thus, in contrast to gene product-related proximal events, where the outcome of germ-line mutations in mice frequently resemble the effect in DT40 cells, alterations of the avian transcriptome may instead predominantly be secondary to clonal events unrelated to the deleted gene.

However, one parameter known to be of importance for DT40 cells is the expression of surface IgM, since the level may change over time [24]. Owing to that, this parameter was not carefully monitored in this study, we can not rule out that this could have influenced our results. Future studies analyzing cell-lines of other origins may shed further light on the underlying mechanism.

Acknowledgments

This work was supported by the Swedish Cancer Foundation, The Swedish Science Council, The Stockholm County Council (research grant ALF-projektmedel medicin) and the FP7 grant EURO-PADnet.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2008.10.040.

References

[1] D. Vetrie, Isolation of the defective gene in X-linked agammaglobulinaemia, J. Med. Genet. 30 (1993) 452–453.
[2] S. Tsukada, D.C. Saffran, D.J. Rawlings, O. Parolini, R.C. Allen, I. Klisak, R.S. Sparkes, H. Kubagawa, T. Mohandas, S. Quan, et al., Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia, Cell 72 (1993) 279–290.
[3] D.J. Rawlings, D.C. Saffran, S. Tsukada, D.A. Largaespada, J.C. Grimaldi, L. Cohen, R.N. Mohr, J.F. Bazan, M. Howard, N.G. Copeland, et al., Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice, Science 261 (1993) 358–361.
[4] J.D. Thomas, P. Sideras, C.I. Smith, I. Vorechovsky, V. Chapman, W.E. Paul, Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes, Science 261 (1993) 355–358.
[5] W. Ellmeier, S. Jung, M.J. Sunshine, F. Hatam, Y. Xu, D. Baltimore, H. Mano, D.R. Littman, Severe B cell deficiency in mice lacking the tec kinase family members Tec and Btk, J. Exp. Med. 192 (2000) 1611–1624.
[6] C.I. Smith, T.C. Islam, P.T. Mattsson, A.J. Mohamed, B.F. Nore, M. Vihinen, The Tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, Tec, Txk and homologs in other species, Bioessays 23 (2001) 436–446.
[7] A.T. Miller, L.J. Berg, New insights into the regulation and functions of Tec family tyrosine kinases in the immune system, Curr. Opin. Immunol. 14 (2002) 331–340.
[8] A. Takesono, L.D. Finkelstein, P.L. Schwartzberg, Beyond calcium: new signaling pathways for Tec family kinases, J. Cell Sci. 115 (2002) 3039–3048.
[9] C.I. Smith, K.B. Islam, I. Vorechovsky, O. Olerup, E. Wallin, H. Rabbani, B. Baskin, L. Hammarstrom, X-linked agammaglobulinemia and other immunoglobulin deficiencies, Immunol. Rev. 138 (1994) 159–183.
[10] C. Ortutay, B.F. Nore, M. Vihinen M, C.I. Smith, Phylogeny of Tec family kinases: Identification of pre-metzoan origin of Btk, Bmx, Itk, Tec, Txk and the Btk regulator SH3BP5, Adv, Genet, 64, in press, doi:10.1016/S0065-2660(08)00803-1.
[11] K. Baba, A. Takeshita, K. Majima, R. Ueda, S. Kondo, N. Juni, D. Yamamoto, The Drosophila Bruton's tyrosine kinase (Btk) homolog is required for adult survival and male genital formation, Mol. Cell. Biol. 19 (1999) 4405–4413.
[12] N. Hamada, C.M. Backesjo, C.I. Smith, D. Yamamoto, Functional replacement of Drosophila Btk29A with human Btk in male genital development and survival, FEBS Lett. 579 (2005) 4131–4137.
[13] M. Takata, T. Kurosaki, A role for Bruton's tyrosine kinase in B cell antigen receptor-mediated activation of phospholipase C-γ2, J. Exp. Med. 184 (1996) 31–40.
[14] X. Peng, C.L. Wood, E.M. Blalock, K.C. Chen, P.W. Landfield, A.J. Stromberg, Statistical implications of pooling RNA samples for microarray experiments, BMC Bioinform. 4 (2003) 26.
[15] C.M. Kendziorski, Y. Zhang, H. Lan, A.D. Attie, The efficiency of pooling mRNA in microarray experiments, Biostatistics 4 (2003) 465–477.
[16] J.H. Shih, A.M. Michalowska, K. Dobbin, Y. Ye, T.H. Qiu, J.E. Green, Effects of pooling mRNA in microarray class comparisons, Bioinformatics 20 (2004) 3318–3325.
[17] C. Li, W.H. Wong, Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection, Proc. Natl. Acad. Sci. USA 98 (2001) 31–36.
[18] J.M. Lindvall, K.E. Blomberg, A. Berglof, C.I. Smith, Distinct gene expression signature in Btk-defective T1 B-cells, Biochem. Biophys. Res. Commun. 346 (2006) 461–469.
[19] J.M. Lindvall, K.E. Blomberg, A. Berglof, Q. Yang, C.I. Smith, T.C. Islam, Gene expression profile of B cells from Xid mice and Btk knockout mice, Eur. J. Immunol. 34 (2004) 1981–1991.
[20] M. Hasan, G. Lopez-Herrera, K.E. Blomberg, J.M. Lindvall, A. Berglof, C.I. Smith, L. Vargas, Defective Toll-like receptor 9-mediated cytokine production in B cells from Bruton's tyrosine kinase-deficient mice, Immunology 123 (2008) 239–249.
[21] M.G. Tomlinson, T. Kurosaki, A.E. Berson, G.H. Fujii, J.A. Johnston, J.B. Bolen, Reconstitution of Btk signaling by the atypical tec family tyrosine kinases Bmx and Txk, J. Biol. Chem. 274 (1999) 13577–13585.
[22] J.B. Petro, W.N. Khan, Phospholipase C-γ2 couples Bruton's tyrosine kinase to the NF-kappaB signaling pathway in B lymphocytes, J. Biol. Chem. 276 (2001) 1715–1719.
[23] Z.K. Mirnics, E. Caudell, Y. Gao, K. Kuwahara, N. Sakaguchi, T. Kurosaki, J. Burnside, K. Mirnics, S.J. Corey, Microarray analysis of Lyn-deficient B cells reveals germinal center-associated nuclear protein and other genes associated with the lymphoid germinal center, J. Immunol. 172 (2004) 4133–4141.
[24] S. Gao, T. Takano, K. Sada, J. He, C. Noda, N. Hori-Tamura, A. Tomoda, H. Yamamura, A novel phenoxazine derivative suppresses surface IgM expression in DT40 B cell line, Br. J. Pharmacol. 137 (2002) 749–755.

OPEN ACCESS Freely available online

PLoS one

Differential Evolutionary Wiring of the Tyrosine Kinase Btk

Hossain M. Nawaz[1], Per Kylsten[2], Noriko Hamada[3], Daisuke Yamamoto[3], C. I. Edvard Smith[1☯], Jessica M. Lindvall [4,5*☯]

1 Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, Huddinge, Stockholm, Sweden, **2** Europaskolan, Strängnäs, Sweden, **3** Division of Neurogenetics, Tohoku University Graduate School of Life Sciences, Sendai, Japan, **4** Bioinformatics and Expression Analysis, Department of Biosciences and Nutrition, Karolinska Institutet, Karolinska University Hospital Huddinge, Huddinge, Stockholm, Sweden, **5** BioinformaticService, Saltsjö-Boo, Sweden

Abstract

Background: A central question within biology is how intracellular signaling pathways are maintained throughout evolution. *Btk29A* is considered to be the fly-homolog of the mammalian Bruton's tyrosine kinase (Btk), which is a non-receptor tyrosine-kinase of the Tec-family. In mammalian cells, there is a single transcript splice-form and the corresponding Btk-protein plays an important role for B-lymphocyte development with alterations within the human *BTK* gene causing the immunodeficiency disease X-linked agammaglobulinemia in man and a related disorder in mice. In contrast, the *Drosophila Btk29A* locus encodes two splice-variants, where the type 2-form is the more related to the mammalian *Btk* gene product displaying more than 80% homology. In *Drosophila*, *Btk29A* displays a dynamic pattern of expression through the embryonic to adult stages. Complete loss-of-function of both splice-forms is lethal, whereas selective absence of the type 2-form reduces the adult lifespan of the fly and causes developmental abnormalities in male genitalia.

Methodology/Principal Findings: Out of 7004–7979 transcripts expressed in the four sample groups, 5587 (70–79%) were found in all four tissues and strains. Here, we investigated the role of *Btk29A* type 2 on a transcriptomic level in larval CNS and adult heads. We used samples either selectively defective in *Btk29A* type 2 ($Btk29A^{ficP}$) or revertant flies with restored *Btk29A* type 2-function ($Btk29A^{(fic\ Exc1-16)}$). The whole transcriptomic profile for the different sample groups revealed Gene Ontology patterns reflecting lifespan abnormalities in adult head neuronal tissue, but not in larvae.

Conclusions: In the *Btk29A* type 2-deficient strains there was no significant overlap between transcriptomic alterations in adult heads and larvae neuronal tissue, respectively. Moreover, there was no significant overlap of the transcriptomic changes between flies and mammals, suggesting that the evolutionary conservation is confined to components of the proximal signaling, whereas the corresponding, downstream transcriptional regulation has been differentially wired.

Citation: Nawaz HM, Kylsten P, Hamada N, Yamamoto D, Smith CIE, et al. (2012) Differential Evolutionary Wiring of the Tyrosine Kinase Btk. PLoS ONE 7(5): e35640. doi:10.1371/journal.pone.0035640

Editor: Edward J. Louis, University of Nottingham, United Kingdom

Received January 18, 2012; **Accepted** March 19, 2012; **Published** May 4, 2012

Funding: This work was supported by the Swedish Cancer Society and the Torsten and Ragnar Söderberg Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have the following interests: J.M. Lindvall is partly employed by BioinformaticService, who provided bioinformatical consultancy help regarding the analyses in this study. J.M. Lindvall is also affiliated with the medical university, Karolinska Institutet. BioinformaticService declares no ownership in the results from the analyses in this work. P. Kylsten is employed by Europaskolan and Europaskolan is an upper secondary school and a college of higher learning and therefore a non-profit organisation. For both BioinformaticService and Europaskolan there are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: jessica.lindvall@ki.se

☯ These authors contributed equally to this work.

Introduction

The evolution of gene expression is considered to mainly result from regulatory, rather than coding, mutations causing phenotypic differences [1]. Analyzing six different organs from ten different species it was recently reported that the rate of gene expression evolution varies among organs, lineages and chromosomes [2]. As gene products commonly function together in distinct combinations to fulfill specific tasks, concerted expression changes of selected genes may be relevant to the survival of the species. Along these lines of arguments, Brawand *et al.* described sets of different organ-specific modules, which were evolutionarily conserved [2]. In this study we have investigated the evolution of tyrosine kinase-based signaling, focusing on Bruton's tyrosine kinase (BTK) in particular. While it is known that elements in proximal BTK-signaling are conserved even among distantly related species, it remains an open question as to whether this is also true for the entire pathways down to the effector level. Here we address this question at the transcriptomic level.

The sequence of Btk has been conserved throughout evolution, with an ancestor emerging already prior to the evolution of metazoans [3]. This kinase belongs to the Tec family of non-receptor tyrosine kinases (TFKs). While insects have only a single TFK, in vertebrates there are several kinase species, which have

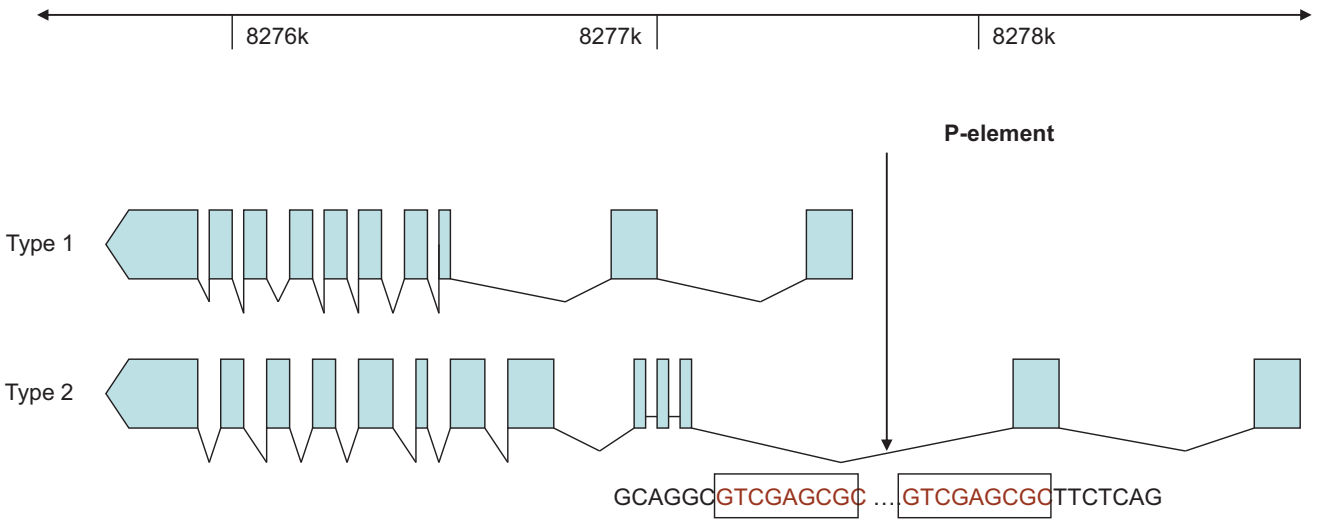

Figure 1. P-element insertion point detection: Schematic figure showing P-element transposon P(BmΔ-w) insertion site. Sequencing results shows P-element insertion site (arrow) within the genomic DNA of *Btk29A* at 2L: 8277721. GTCGAGCGC repeats can be seen at both end of the P insertion, which is a characteristic mark of transposon insertion. This information was further use to distinguish between mutant and revertant flies. In mutant flies this breakpoint was easily detected in amplicons (Materials and Methods), whereas the revertant did not show any P insertion.
doi:10.1371/journal.pone.0035640.g001

evolved through gene duplications. The fly kinase is most homologous to vertebrate Btk. However, in spite of the high degree of sequence conservation, the functional role of Btk seems to vary throughout evolution. In higher organisms, such as humans and other mammals, the significance of Btk lies in its function for a normal development of the immune system [4].

In the absence of mammalian Btk, B-cell receptor signaling is insufficient for the generation of mature B-lymphocytes [5,6,7], resulting in the immunodeficiency disease X-linked agammaglobulinemia (XLA) in humans [8,9] and X-linked immunodeficiency disease (*Xid*) in mice [10,11]. Insects, like *Drosophila*, possess neither B- nor T-cells. An orthologous function for *Drosophila* Btk, i.e., regulating B-cell maturation, can therefore not be expected. The *Drosophila Btk29A* locus produces two different gene products, type 1 and type 2, respectively, by differential splicing. The type 2 form reveals the highest homology to human BTK among mammalian TFKs [12]. Thus, this variant is considered to be the fly homolog of Btk by means of protein sequence [3,13]. It is specifically required for longevity and for development of male genitalia in the fly [12].

The type 1 splice variant is shorter at the N-terminus and is unique to flies [3,12]. The *Drosophila Btk29A* locus displays a dynamic pattern of expression through the embryonic to adult stages [FlyAtlas. http://130.209.54.32/atlas/atlas.cgi [14]. The $Btk29A^{ficP}$ is a unique allele in that it is devoid of transcription of the type 2 isoform, while leaving the type 1 isoform intact. *Btk29A* types 1 and 2 are both expressed in the central nervous system (CNS) and in the imaginal discs [12], which are epidermal thickenings in the larvae containing ecto- and mesodermal cells, which give rise to the adult organs during metamorphosis. Complete loss of function of the gene (i.e., loss of both types 1 and 2) in female germline cells, produced by using the dominant-female-sterile, FLP/FRT technique, results in oocyte undergrowth and subsequent embryonic death accompanied by defective head involution [15,16,17,18,19]. Offspring with selective loss of the type 2 transcript are viable, developing malformed male genitalia and a reduced adult life span [12]. Thus, the *Btk29A* locus exerts pleiotropic functions both through distinct spatio-temporal riming of expression as well as the generation of distinct forms of protein products by alternative splicing in various tissues. When *Btk29A* function is lost in a *Src64* mutant background, cellularization becomes incomplete in the blastoderm-stage embryo [20] and late-staged embryos fail to complete dorsal closure [21]. In *Btk29A* mutant females, oogenesis is underdeveloped presumably due to deficits in the formation of ring canals that transfer cytoplasm from nurse cells to oocytes [15,16,22]. Both the cellularization and oocyte phenotypes appear to result from failure to activate actin–myosin contractions [20]. Chandrasekaran and Beckendorf *et al.* have shown that *Btk29A* controls both the actin cytoskeleton and the cell cycle in the morphogenesis of embryonic salivary glands [23]. Interactions between mammalian Btk and actin have also been reported in several settings [24,25,26,27,28], suggesting this to be a common denominator in the proximal part of the Btk-signaling pathway, i.e. proximal of the Btk-dependent transcriptional regulation.

In the present study we adopted a genome-wide approach to identify Btk-dependent targets in neuronal tissues by exploring the transcriptional output from Btk-deficient and wild-type tissues, for two developmental stages in *Drosophila melanogaster*, respectively. Genes identified in this way could be *direct* or *indirect* targets for Btk-regulated transcription and outline part of the transcriptomic role of Btk in the development of the fly. The identification of Btk targets, corroborated by statistical analyses and gene set enrichment analyses, reveals parts of the scope and complexity, which Btk plays in the fly. We also conclude that there is no significant functional transcriptomic conservation for Btk targets between mouse B-cells and neuronal tissue from *Drosophila*.

Results and Discussion

We performed transcriptional profiling of the central nervous system (CNS) tissues from mutant ($Btk29A^{ficP}$) and revertant ($Btk29A^{fic\ exe1-16}$) adults and larvae using the Affymetrix *Drosophila* Genome 2 chips, with 18,880 probe sets covering around 13,500 genes. This analysis yielded a list of affected genes known to function in longevity and aging, two biological processes impaired in $Btk29A^{ficP}$ mutants, thus validating the experimental rationale and setup. Although the phenotype-genotype association in the

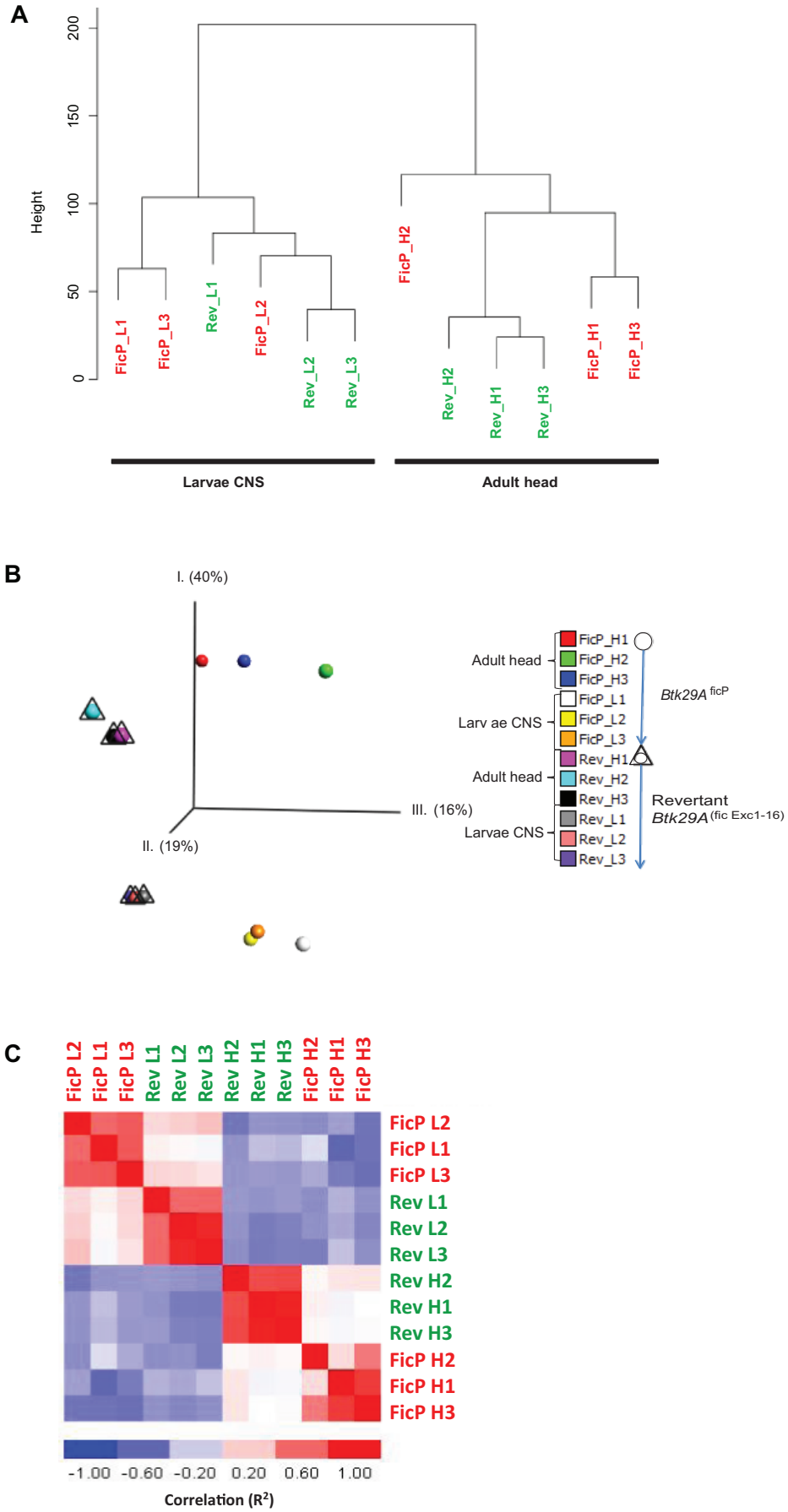
A
Height
0
50
100
150
200
FicP_L1
FicP_L3
Rev_L1
FicP_L2
Rev_L2
Rev_L3
FicP_H2
Rev_H2
Rev_H1
Rev_H3
FicP_H1
FicP_H3
Larvae CNS
Adult head
B
I. (40%)
II. (19%)
III. (16%)
Adult head
Larv ae CNS
Adult head
Larvae CNS
Btk29A ficP
Revertant
Btk29A (fic Exc1-16)
C
FicP L2
FicP L1
FicP L3
Rev L1
Rev L2
Rev L3
Rev H2
Rev H1
Rev H3
FicP H2
FicP H1
FicP H3
-1.00 -0.60 -0.20 0.20 0.60 1.00
Correlation (R2)

Figure 2. Global patterns of gene expression differences among sample groups: ***A) Hierarchical clustering:*** Neighbor joining tree based on pairwise distance matrices (1−ρ, Spearman correlation coefficient) for the different sample groups within this study. Sample groups fall out based on the strongest factor, larval CNS or adult head, respectively, i.e. developmental stage. Thereafter on either *Btk29A*^{ficP} (red) or revertant (*Btk29A*$^{(fic\ Exc1-16)}$) (green). ***B) Factorial map of the principal-component analysis:*** The replicates within each sample collection are grouped together based on the individual samples whole genome expression profile. Revertant is denoted with Δ and ficP with a o. All samples are colored individually. The proportion of the variance explained by the principal components (axis) is indicated in parentheses in the graph. ***C) Sample correlation matrix:*** Spearman correlation (R^2) is calculated and visualized by color (red-blue) in the matrix. Within the replicates for the individual sample groups the correlation is higher than between the sample groups. Also, the correlation within developmental stage was found to be higher compared to between mutant strains. Sample names are color-coded with red (*Btk29A*^{ficP}) and green (revertant).
doi:10.1371/journal.pone.0035640.g002

Drosophila Btk29A locus has been studied in some detail, the components and regulators of Btk29A signaling remain unexplored on the global transcriptomic level. Also, the biological processes triggered by Btk signaling defects in *Drosophila* are less well understood. Moreover, to our knowledge, the comparison of changes in gene expression profiles between Btk mutants of different animal species has not been performed before. Here, we use the strength of the *Drosophila* system in order to identify candidate effectors that take part in the Btk signaling process and use these data to perform an inter-species comparison of Btk-dependent components between mouse and fruit fly. To achieve this we have made use of the *Btk29A* type 2 mutant (*Btk29A*^{ficP}) and a revertant strain (*Btk29A*$^{(fic\ Exc1-16)}$) where wild type gene function has been restored by a jump-out event of the mutagenic P-element from the *Btk29A*^{ficP} chromosome. This revertant fly is considered to be the most accurate wild type control for the mutant as wild-type development and life expectancy are fully restored and with the exception of the P-element, the *Btk29A*-carrying chromosome is the same for the two strains [13].

Verifying P-element Insertion Point and the Nature of the Reversion

In order to define molecularly the experimental flies, we initially determined the exact location of the P-element insertion. For this we made use of P-element-specific primers (directed outwards from both the 5′ and 3′-ends of the P-element) for *Btk29A*^{ficP} and eight primers ~500 bp apart, covering in total a stretch of 4 kb along the genomic sequence of the *Btk29A* locus. Using PCR, amplicons were detected for both ends of the P-element. The PCR products were sequenced and the results showed that the P-element point-of-insertion into the genomic sequence was at nucleotide 2L: 8,277,721 (Figure 1). This indicates that the P-element sits within the *Btk29A* locus, although its location deviates for 866 bp from that given in FlyBase (http://flybase.org/) showing 2L:8,276,855.8,280,039 [-]. The relevant genomic primer pair produced a PCR-product from *Btk29A* $^{fic\ exe\ 1-16}$ revertant genomic DNA, which upon sequencing showed the P-element to having left the locus by perfect excision, leaving only the wild-type genomic sequence (data not shown).

Transcriptional Profiling

The transcriptional profiles were analyzed with Affymetrix whole genome arrays (GeneChip *Drosophila* Genome 2.0) by a comparative approach between the "mutant" and "revertant" sample groups for either larvae or adult heads. Thus, in total, 4 different sample groups were collected (Figure 2A–C). To reconstruct strain and tissue trends in a global transcriptomic detail, we built an expression distance matrix for the four sample groups with its replicas and reconstructed a gene expression tree (Figure 2A). The tree is highly consistent with the expectation that

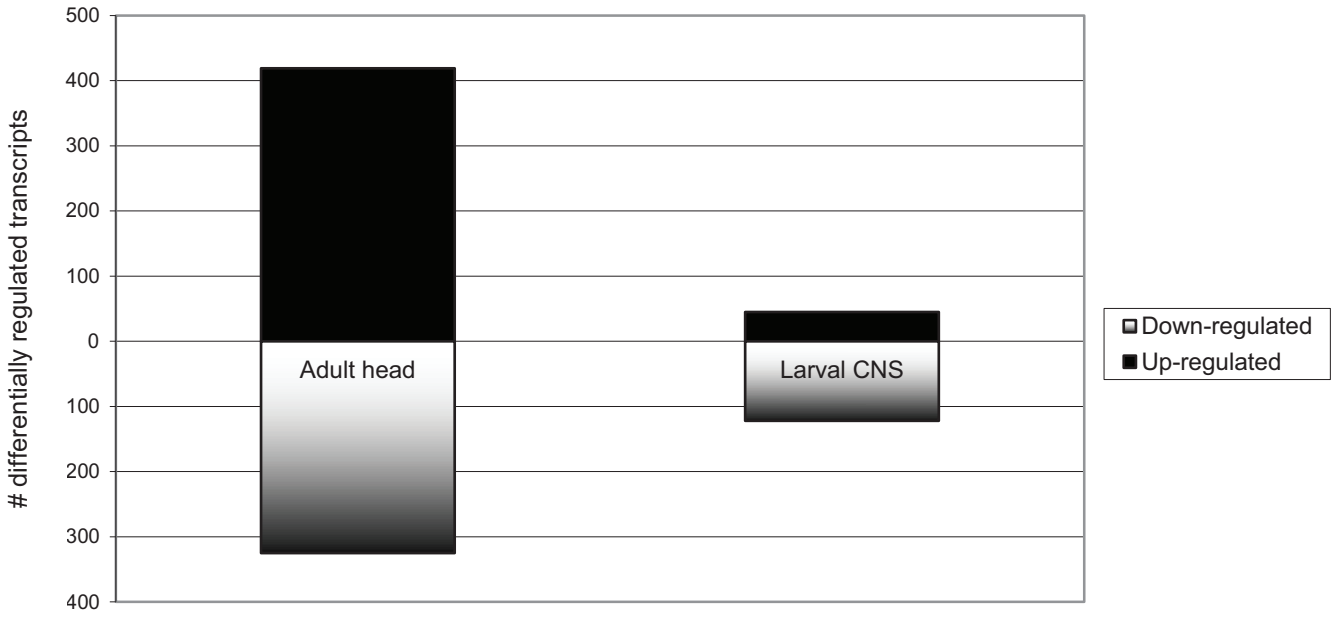

Figure 3. Number of differentially expressed transcripts in Btk defective flies: A differential fold-change cut-off of the Signal Log Ratio (SLR) >1,2 (difference of means between *Btk29A*^{ficP} and revertant (*Btk29A*$^{(fic\ Exc1-16)}$) was applied to define genes whose expression was significantly different from that of the revertant and *Btk29A*^{ficP}. Bar-graph showing the number of up- and down-regulated transcripts, respectively, in adult head and larval CNS after the pair-wise comparison of revertant (*Btk29A*$^{(fic\ Exc1-16)}$) and *Btk29A*^{ficP} data.
doi:10.1371/journal.pone.0035640.g003

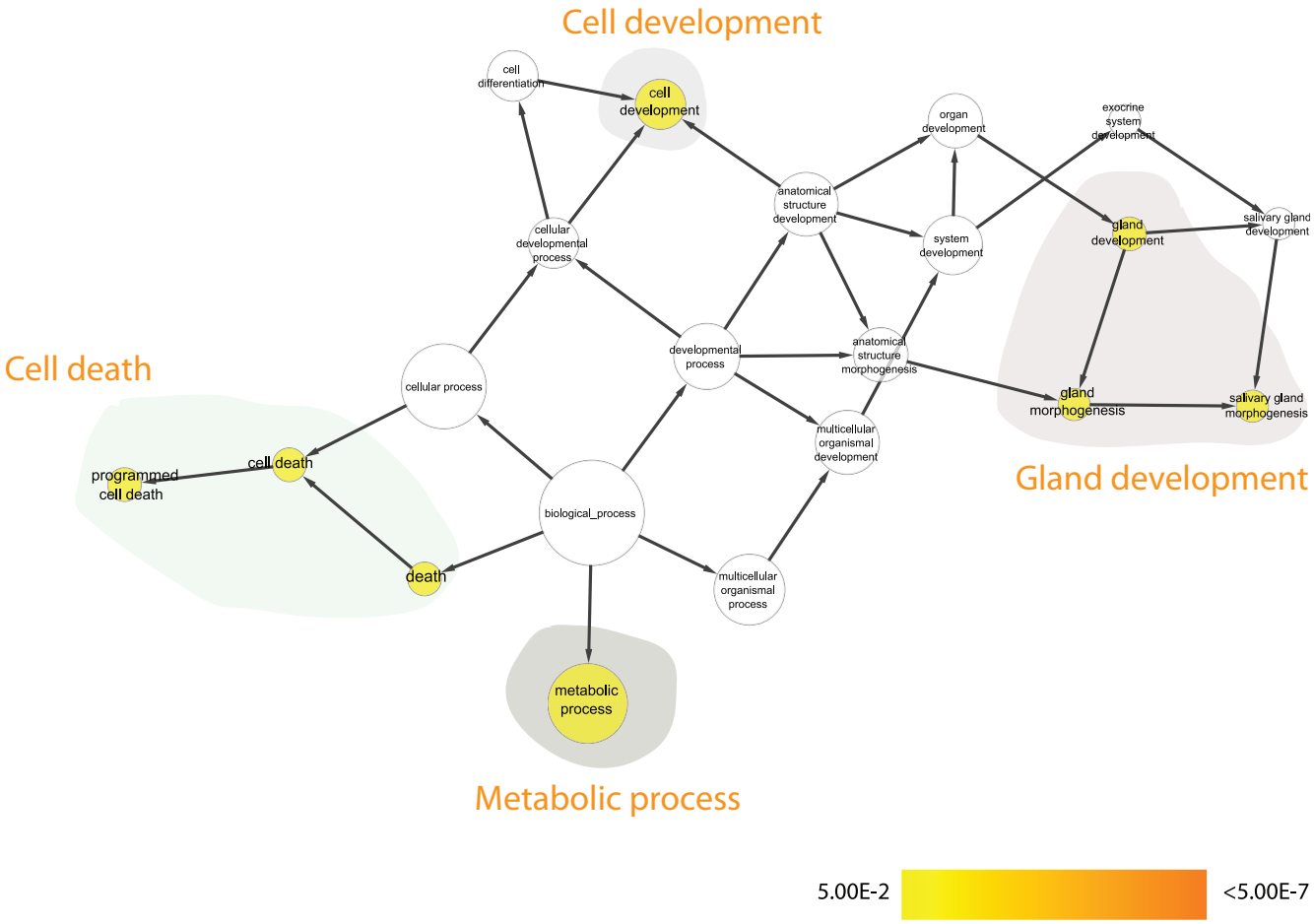

Figure 4. GSEA for the 391 Btk-dependent transcripts during Fly neuronal development: Enriched Biological Process clusters within the list of 391-transcripts (Btk-dependent transcripts during fly neuronal development). Figure 4 should be statistically interpreted as follows: The nodes, corresponding to different Gene Ontology clusters, are either not colored (white) i.e. not found with statistical power or colored in the scale yellow to orange, where yellow nodes are found with statistical significance *after* Bonferroni correction $p<0.05$ and orange colored nodes are found to be even more statistically significant after correction, with a $p<7*10^{-8}$. The grey-zoned data highlights statistically enriched clusters of nodes (genes/transcripts), which all are represented under the manually designated heading e.g. 'Gland development' or 'Cell death'. Due to space limitations in the main figure (Figure 4) we are not able to list the genes belonging to each grey-zone and cluster. This information is instead found in the Figure S1.

doi:10.1371/journal.pone.0035640.g004

the predominant factor to characterize the profile is the tissue type/developmental stage followed by the Btk-genotype. A quality measure for the data input is that the majority of replicates fall within the respective sample group (Figure 2A–C). To obtain an initial overview of the transcriptional expression patterns, we performed a principal-component analysis, which clearly separates the data according to sample group (Figure 2B). Figure 2C represents a Pearson correlation (R^2) matrix for the whole transcriptomic profile for all samples included in the study. Here we see a higher intra-tissue correlation between revertant and $Btk29A^{ficP}$ mutant of the same developmental stage compared to the intra-strain correlation between the two stages/tissues (Figure 2C).

Under such circumstances, several mathematical approaches are possible in extracting the genes that behave differently according to sample groups. When applying an ANOVA filtering in the comparison between the mutant ($Btk29A^{ficP}$) and revertant ($Btk29A^{(fic\ Exc1-16)}$) results, irrespective of the stage/tissue (larvae CNS or adult head) using a *p*-value of 0.05, we found 523 transcripts being statistically different between the $Btk29A^{ficP}$ and the revertant. On the other hand, when considering the stage/tissue as the decisive factor (regardless of using $Btk29A^{ficP}$ or revertant data) we detected 4489 transcripts being statistically different between the groups. This indicates, as expected, that the difference between tissue types or developmental stages in the fly gives a stronger influence on the transcriptome compared to the influence of the $Btk29A^{ficP}$ mutation. On the other hand, when *both* the tissue type (larvae CNS and adult head) *and* genotype ($Btk29A^{ficP}$ and revertant) are considered as the decisive ANOVA factors with a *p*-value <0.05 we find 391 transcripts being statistically different between the 4 sample groups. Thus, on a transcriptomic level there are 391 transcripts that, by these criteria, are Btk-dependent in *Drosophila* neuronal tissue development, from larvae to adult flies.

A differential fold-change cut-off of Signal Log Ratio (SLR) >1,2 (difference of means between FicP and revertant) was applied to define genes whose expression was significantly different between the revertant and $Btk29A^{ficP}$. The number of transcripts found to be differentially expressed between the two genotypes was more than 4 times higher in adult heads (744) as compared to

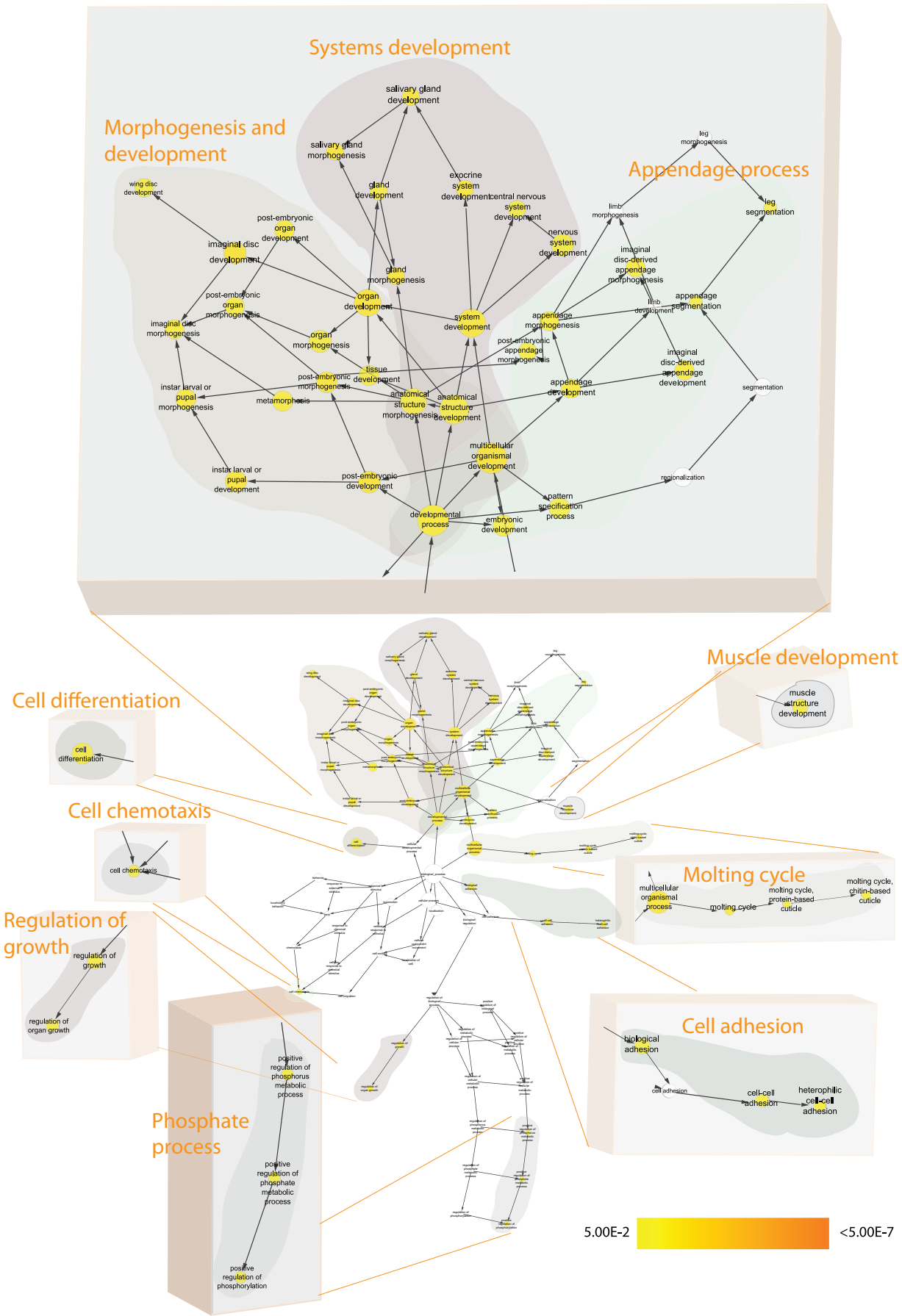
Systems development
Morphogenesis and development
Appendage process
Muscle development
Cell differentiation
Cell chemotaxis
Regulation of growth
Molting cycle
Cell adhesion
Phosphate process
salivary gland development
salivary gland morphogenesis
wing disc development
gland development
exocrine system development
central nervous system development
leg morphogenesis
leg segmentation
limb morphogenesis
post-embryonic organ development
imaginal disc development
nervous system development
imaginal disc-derived appendage morphogenesis
gland morphogenesis
organ development
post-embryonic organ morphogenesis
system development
appendage morphogenesis
limb development
appendage segmentation
imaginal disc morphogenesis
organ morphogenesis
post-embryonic appendage morphogenesis
imaginal disc-derived appendage development
tissue development
post-embryonic morphogenesis
appendage development
segmentation
instar larval or pupal morphogenesis
metamorphosis
anatomical structure morphogenesis
anatomical structure development
multicellular organismal development
instar larval or pupal development
post-embryonic development
regionalization
pattern specification process
developmental process
embryonic development
muscle structure development
cell differentiation
cell chemotaxis
regulation of growth
regulation of organ growth
multicellular organismal process
molting cycle
molting cycle, protein-based cuticle
molting cycle, chitin-based cuticle
biological adhesion
cell adhesion
cell-cell adhesion
heterophilic cell-cell adhesion
positive regulation of phosphorus metabolic process
positive regulation of phosphate metabolic process
positive regulation of phosphorylation
5.00E-2
<5.00E-7

Figure 5. GSEA for the 167 Btk-dependent transcripts found in larvae CNS: Differentially expressed transcripts (167) were subjected to GSEA and enriched clusters were found. Figure 5 should be statistically interpreted as follows: The nodes, corresponding to different Gene Ontology clusters, are either not colored (white) i.e. not found with statistical power or colored in the scale yellow to orange, where yellow nodes are found with statistical significance *after* Bonferroni correction $p<0.05$ and orange colored nodes are found to be even more statistically significant after correction, with a $p<7*10^{-8}$. The grey-zoned data highlights statistically enriched clusters of nodes (genes/transcripts), which all are represented under the manually designated heading e.g. 'Systems development' or 'Regulation of growth'. Due to space limitations in the main figure (Figure 5) we are not able to list the genes belonging to each grey-zone and cluster. This information is instead found in the Figure S2.
doi:10.1371/journal.pone.0035640.g005

larval CNS (167; Figure 3). This suggests that there are more Btk-dependent transcripts in the head, perhaps also reflecting the fact that the head is not only composed of neuronal tissue. The distribution of up- and down-regulated transcripts was approximately 50% in adult heads. In the larval CNS sample group, the distribution of differentially expressed genes/transcripts was approximately 30% up-regulated versus 70% down-regulated, suggesting that transcriptional loss-of-function is the predominating feature in the larval CNS of *Btk29A*^{ficP} mutants.

Clustering of Genes and Functional Enrichment

By performing gene set enrichment analysis using Cytoscape and the plug-in BiNGO we identified different Gene Ontology (GO) clusters being enriched in the different data sets. For the 391 probe sets indicated to be *Btk-dependent during fly neuronal development* we found four major Gene Ontology clusters to be enriched within this list (Figure 4 and a more detailed view is found in Figure S1 where the corresponding genes are listed to respective statistically significant nodes). Due to space limitations in the figures we were not able to enlarge all the titles of the nodes within the figure. We have instead tried to find commonalities within the grey-zone and manually put a 'heading' for each of these zones in order to give the reader an overview of the result of the GSEA. In order to see the node titles (corresponding to Gene Ontology Biological Process names) the reader can zoom in on the figure and by this be able to read the text. Amongst these clusters representing Death, Cell development, Metabolic process and Gland development we find genes previously linked to Btk function and fly development, thus validating the approach of our study, but also genes not previously known to be associated with Btk in the fly. As such, gene set enrichment analysis suggests that the genes identified reflect a *bona fide* response of fly neuronal development to the loss of Btk.

In the larval CNS we find 167 Btk-dependent transcripts being differentially expressed with an SLR >1.2 between *Btk29A*^{ficP} and revertant flies. A Gene Set Enrichment Analysis on this set of transcripts reveals Gene Ontology terms which mirror undifferentiated progenitor cells for the future adult i.e., 'imaginal' cells (Figure 5 and a more detailed view is found in Figure S2 where the corresponding genes are listed and colored depending on the direction of the regulation to respective statistically significant nodes). Indeed, one of the processes showing the highest statistical score (after correction with the Benjamini-Hochberg algorithm) is 'Imaginal disc development' (GO:0007444'') with a corrected p-value <0.003. Another GO-term, 'Developmental process' (for *D. melanogaster*), characterized by 42 genes (in our 167 gene list) out of a total of 117(in the GO 'Developmental process' gene list), which comprises 35.8% of the total number of genes in the GO term and 25% of the genes being differentially expressed in ''larval CNS *Btk29A*^{ficP}''.

For the 744 differentially expressed (SLR>1.2) transcripts found in ''adult head *Btk29A*^{ficP}'' there is an overrepresentation of the terms 'Aging' (GO:0007568) and 'Determination of lifespan' (GO:0008340) (Figure 6 and a more detailed view is found in Figure S3 where the corresponding genes are listed and colored depending on the direction of the regulation to respective statistically significant nodes). These findings were expected as life span is reduced in *Btk29A*^{ficP} mutants [12,13]. As these Gene Ontology terms are not found to be enriched in the larval CNS tissue samples we hypothesize that the reduction of life span, due to the Btk defect, is manifested later in the life of the fly and therefore cannot be foreseen at the larval stage. There are no Gene Ontology terms found to be enriched in *both* adult heads and larval CNS from a global transcriptomical point of view. This potentially indicates that Btk plays somewhat different roles at these two developmental stages in *Drosophila*. This is further supported by the finding that the larval CNS and adult head sample groups shared only 20 transcripts (corresponding to 20 genes) out of a total of 29 transcripts (corresponding to 25 genes), whose expression was either up- or down-regulated in *Btk29A*^{ficP} mutants at *both* developmental stages, and a half of these genes was up-regulated and another half was down-regulated (Table 1 and Table S1). Table S1 mirrors Table 1 in addition to adding the dimension of every gene's Gene Ontology term including Gene Ontology ID. Btk might be more important in adult rather than larval neural tissues due to the finding of 'Spindle organization'-enriched genes, which belong to the Gene Ontology term 'mitotic spindle'. Thus, in the adult developmental stage, Btk might play a role in proliferation of the glia, since there are no neuroblasts in the adult head. On the other hand, a gene set enrichment analysis (using the web-based GO tool 'DAVID') performed on the above mentioned 25 genes (Table 1 and Table S1) revealed a statistical significant overrepresentation of the GO term 'Behavior' (GO:0007610), indicating that the Btk defect, independent of the developmental stage, might be manifested via the biological process 'Behavior' in *Drosophila*. Another interesting observation throughout the gene set enrichment analysis was that the number of down-regulated transcripts predominates independent of data set introduced, mirroring the loss-of-function nature of *Btk29A*^{ficP} mutants.

Transcriptional Comparison between Mouse and Fruit fly Btk-defective Cells

A central question in biology is to what level protein function in intracellular signaling pathways is conserved through evolution of species, like e.g. between mammals and insects. For instance, it has been proven possible to 'humanize' the fly by introducing human genes of interest, including the human *BTK* gene, and studying them in an organotypic context [13,29,30]. Previous studies on components in the JAK/STAT signaling pathway has revealed a small, but statistically significant, overlap between *Drosophila* and mammals at a transcriptomic level [31]. In sea urchins and sea stars, organisms that diverged from their common ancestor 500 million years ago, a three-gene feedback loop involving Notch-signaling controls endoderm and mesoderm development in both overlapping and distinct ways [32]. Furthermore, appendages of different insects show divergent use of developmental regulatory genes, including the helix-loop-helix, homeodomain transcription factor Distal-less [33]. These phenomena have been referred to as gene regulatory network ''plug-ins'', in which sub-circuits are frequently re-deployed during evolution while the internal

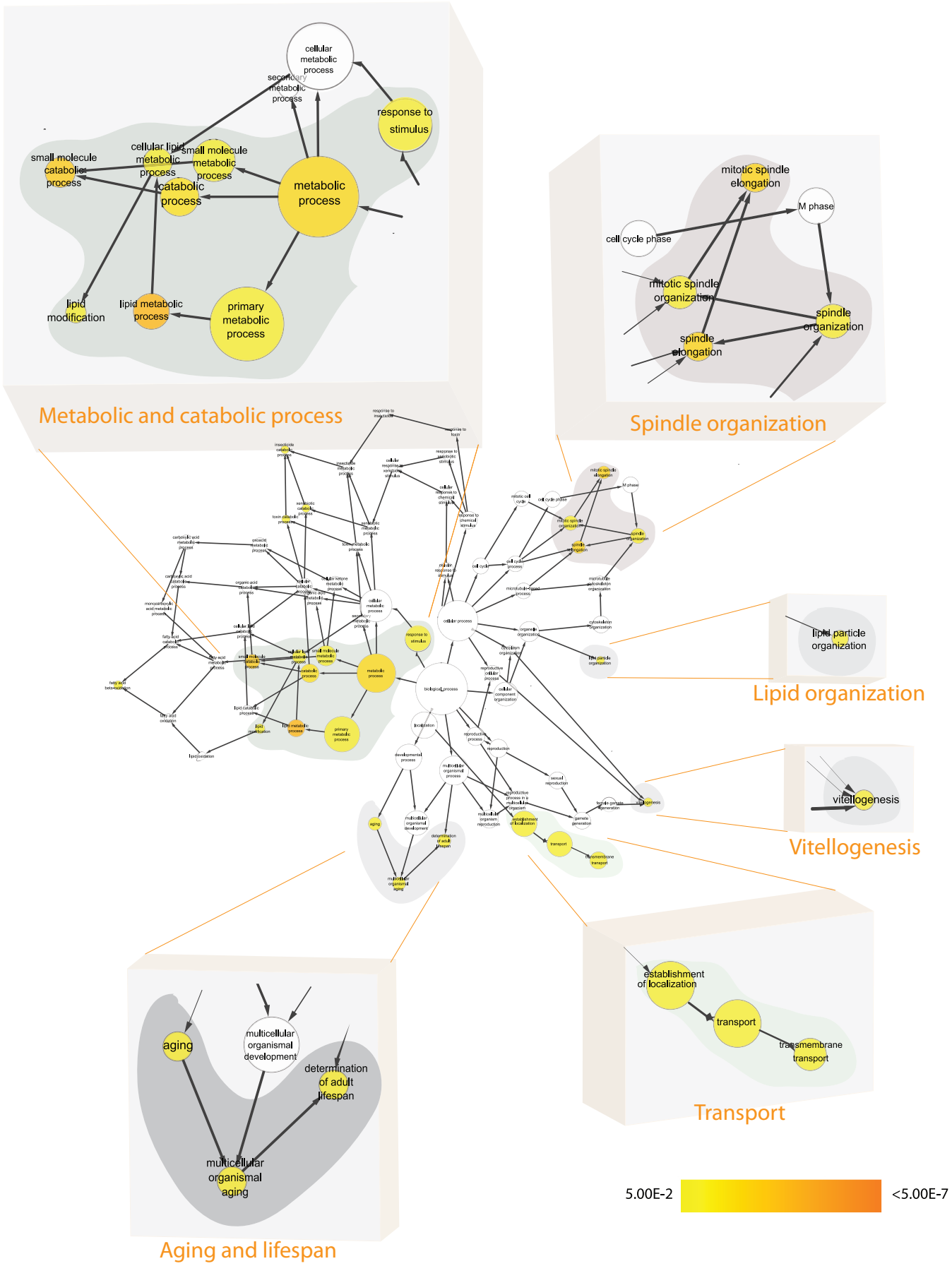
cellular metabolic process
secondary metabolic process
response to stimulus
small molecule catabolic process
cellular lipid metabolic process
small molecule metabolic process
catabolic process
metabolic process
lipid modification
lipid metabolic process
primary metabolic process
Metabolic and catabolic process
mitotic spindle elongation
M phase
cell cycle phase
mitotic spindle organization
spindle organization
spindle elongation
Spindle organization
lipid particle organization
Lipid organization
vitellogenesis
Vitellogenesis
establishment of localization
transport
transmembrane transport
Transport
aging
multicellular organismal development
determination of adult lifespan
multicellular organismal aging
Aging and lifespan
5.00E-2
<5.00E-7

Figure 6. GSEA for the 744 Btk-dependent transcripts found in adult head tissue: Enriched clusters of Biological Processes were found in the Btk-dependent Adult head tissue analysis and were statistically significant. Figure 6 should be statistically interpreted as follows: The nodes, corresponding to different Gene Ontology clusters, are either not colored (white) i.e. not found with statistical power or colored in the scale yellow to orange, where yellow nodes are found with statistical significance *after* Bonferroni correction $p<0.05$ and orange colored nodes are found to be even more statistically significant after correction, with a $p<7*10^{-8}$. The grey-zoned data highlights statistically enriched clusters of nodes (genes/transcripts), which all are represented under the manually designated heading e.g. 'Transport' or 'Aging and life span'. Due to space limitations in the main figure (Figure 6) we are not able to list the genes belonging to each grey-zone and cluster. This information is instead found in the Figure S3.
doi:10.1371/journal.pone.0035640.g006

structure remains the same [34]. Further example of such rewiring comes from protein kinase A (PKA) catalytic subunit signaling in the fungus of the genus, *Cryptococcus* [35]; two sibling species of this pathogen express two different catalytic subunits of PKA, and alternative subunits are used in virulence factor production and mating in each species. It is envisaged that an ancestral PKA underwent a duplication event leading to the two catalytic subunit genes, one of which retained its function for the given biological processes in each species. Whether the ''unused'' subunit has undergone neofunctionalization with a novel gain-of-function for another biological activity is not known, but this example demonstrates evolutionary reconfiguration of a signaling cascade.

Table 1. 25 genes found to overlap between *D. melanogaster* larvae CNS and adult head.

D.melanogaster	FlyBase ID	Differentially regulated	
Gene symbol		***Btk29A*^{ficP}**	***Btk29A*^{ficP}**
		Adult Head	**Larvae CNS**
w	FBgn0003996	4.37	3.49
TpnC47D	FBgn0010423	2.22	2.89
A-Est1	FBgn0015568	1.83	2.21
l(3)mbn	FBgn0002440	1.39	2.18
CG5597	FBgn0034920	1.36	2.12
CG5023	FBgn0038774	1.28	1.62
CG11807	FBgn0033996	2.92	1.59
CG4398	FBgn0034126	1.91	1.53
pnt	FBgn0003118	1.4	1.31
CG2177	FBgn0039902	1.44	1.25
mthl3	FBgn0028956	*−2.23*	*−1.22*
pen-2	FBgn0053198	*−1.84*	*−1.25*
gdl-ORF39	FBgn0028377	*−1.23*	*−1.4*
CG14033	FBgn0046776	*−2.05*	*−1.4*
pst	FBgn0035770	*−1.59*	*−1.56*
CG6984	FBgn0034191	*−1.42*	*−1.65*
CG11671	FBgn0037562	*−2.01*	*−2*
CG42254	FBgn0259112	*−1.79*	*−2.02*
CG17264	FBgn0031490	*−1.72*	*−2.07*
CG32368	FBgn0052368	*−2.9*	*−5.88*
CG12241	FBgn0038304	1.46	*−1.29*
Dob	FBgn0030607	*−1.87*	1.61
Obp56h	FBgn0034475	*−3.52*	2.12
proPO-A1	FBgn0261362	*−1.64*	2.81
CG10176	FBgn0032682	1.22	*−1.28*

Italics denotes down-regulated genes.
*denotes genes differentially expressed NOT in the same regulatory direction for the larval CNS and adult head sample group.
doi:10.1371/journal.pone.0035640.t001

Recently, rewiring of both prokaryotic and eukaryotic signaling pathways has been achieved using rational design, demonstrating another aspect of the alteration of signal transduction pathways [36,37].

Apart from the well-known developmental role in the immune system, mammalian Btk has been shown to exert two counteracting roles in apoptosis, one as a protector and in the other as an inducer of apoptosis depending on the context [38,39]. This reflects the diverse role of a protein within a species. In the fruit fly, loss of the *Btk29A* type 2 transcript is compatible with life, as opposed to loss of both types 1 and 2 of *Btk29A*, which is embryonic lethal. However, type 2 mutant flies display reduced life span as well as malformation of the male genitalia [15,16]. We have previously published work on Btk-defective mouse B-lymphocytes using gene expression profiling [40,41]. In order to identify factors with conserved functions throughout evolution we conducted an inter-species comparison of the Btk-dependent transcripts identified in *Drosophila* against our previous transcriptomic data obtained with mouse Btk-defective Transitional type 1 B-lymphocytes [40]. Figure 7 illustrates the number of transcripts found to be expressed in every *Drosophila* sample group examined and the level of overlap between the assemblies. This indicates that approximately 30% of the *Drosophila* genome is expressed at the time examined in the respective tissues and strains. In comparison to these numbers, in mammalian B-lymphocytes [40] we see that 37% of all transcripts in the mouse genome are expressed at any given time. Figure 8A shows a Venn-diagram that illustrates the overlap between the differentially expressed transcripts found in mouse Btk-defective Transitional type 1 B-cells (a total of 147 regulated genes) and the *Drosophila* *Btk29A*^{ficP} adult head (a total of 744 differentially expressed transcripts). The overlap between the two species is only sixteen transcripts, corresponding to 13 genes in the Btk-defective mice found in our previous study [40] having orthologs in the Btk-dependent transcripts identified in *Drosophila* (Figure 8B). Of these 16 transcripts only 5 are found to show parallel changes in *Drosophila* and mice being either up- or down-regulated in the Btk-defective strains (denoted as * in Figure 8B). By analyzing the gene expression profile from Btk-defective flies representing two different developmental stages and comparing these to mammalian Btk-defective B-cells we conclude that there is no significant overlap in the transcriptome for Btk-defective mammalian B-cells and neuronal cells from *Drosophila*. Based on these observations, we conclude that there is no significant functional transcriptomic conservation for Btk targets between the mammals and fly species.

Concluding Remarks

Although the upstream signaling protein components of *Btk29A* seem to be conserved throughout evolution, the downstream transcriptional pattern seems not to be comparable between the fruit flies and mice. The Btk-dependent gene expression profile seen in mouse transitional type 1 B-lymphocytes from Btk-defective animals thus differs from the global transcriptomic signature seen in *Btk29A* type 2-defective neural tissues from *Drosophila*. This is in contrast to JAK/STAT signaling in which

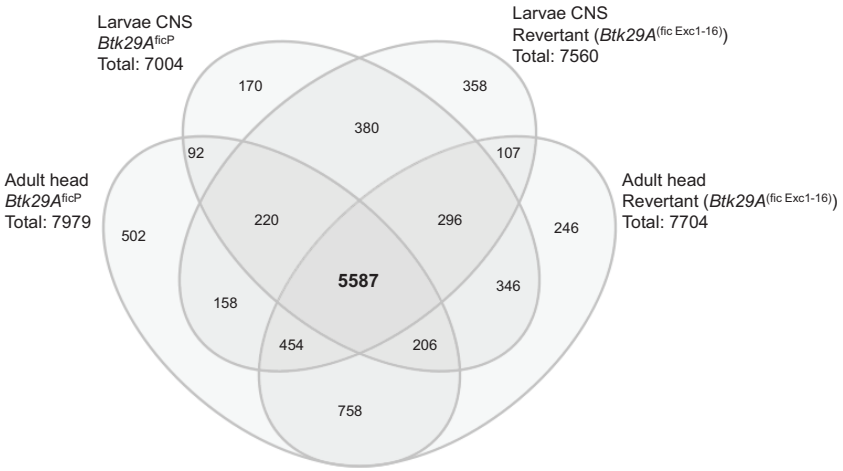

Figure 7. Distribution of expressed transcripts in each sample group and their overlaps: Venn- diagram showing the number of expressed probe-sets of the respective sample group and their overlap. The probe-sets found to be expressed above background (Affymetrix P-calls) in all three replicates per sample group were considered within this figure.
doi:10.1371/journal.pone.0035640.g007

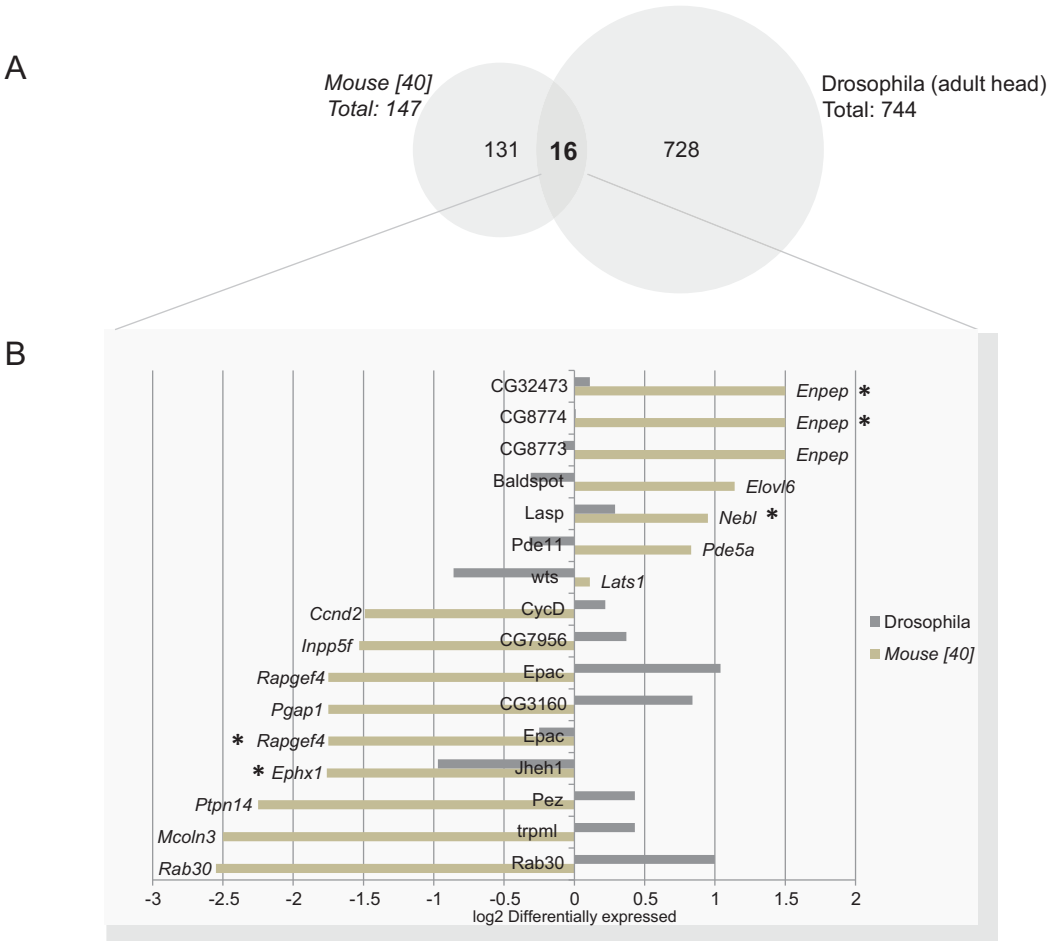

Figure 8. Comparison between Mouse and Drosophila Btk-dependent transcript: A) ***Venn-diagram:*** showing the overlap between Btk-dependent Transitional type 1 B-cells from Btk-defective mice [40] (a total of 147 differentially expressed genes) and the Btk-dependent transcripts found in *Btk29A*^{ficP} *Drosophila* adult head (a total of 744 differentially expressed genes). **B)** ***Thirteen orthologous transcripts found in Btk KO mouse Transitional type 1 B-lymphocytes and in Btk29A defective flies.*** Bar-graph showing the 16 transcripts, corresponding to 13 genes, found to be common in a homology search between the Btk-dependent transcripts found in the *Btk KO* mouse Transitional type 1 B-cells compared to its healthy control strain [40] and the current *Drosophila Btk29A*^{ficP} study. Out of these 13 genes, 5 genes showed the same regulatory direction i.e. being up- or down-regulated in the respective Btk-defective strain (denoted as *).
doi:10.1371/journal.pone.0035640.g008

both the upstream and downstream components were reported to be conserved [31].

Indeed, large scale profiling data must be interpreted with caution and the genes identified here await ultimate proof as to whether they represent the *bona fide* effectors of *Btk29A*-mediated developmental signaling. Although detailed mechanisms of action of individual effectors and their roles linked to *Btk29A* function remains partially unknown, it is interesting to note that a profile related to life-span was recognized, suggesting that our transcriptomical mapping approach has effectively identified different pathways and effectors likely to play roles in Btk signaling and functioning regarding fruit fly development.

Materials and Methods

Drosophila Melanogaster *Strains*

w;Btk29A^{ficP}*/CyO* and *w1118; Btk29A*$^{(fic\ Exc1-16)}$*/SM1* were generated in the Yamamoto laboratory (http://www.jst.go.jp/erato/project/yks_P/yks_P.html). Flies were raised on standard medium on a 12:12 h L:D cycle, at 23°C and at 55% RH. The *Btk29A*^{ficP} chromosome was put over the *CyO, P{w[+mC] = ActGFP}JMR1* (source: Bloomington stock center) balancer using standard crossing schemes. This balancer was later used for sorting heterozygous (GFP-positive) from homozygous *Btk29A*^{ficP} mutants (GFP-negative). *Btk29A*$^{(fic\ Exc1-16)}$ flies were kept homozygous viable in stocks.

P-element Breakpoint Determination using PCR

Four kb (2L:8274950,8279050) surrounding the *Btk29A*-locus was used as a template to construct 8 forward-, and 8 reverse-oriented primers covering the entire 4-kb region from both ends with a 500 bp spacing. Primers were also made for the 3′- and 5′-ends of the *Btk29A*^{ficP} -responsible P-element (BmΔ-w). Both primers were facing outwards from the P-element. PCR was performed using ABI GeneAmp^{TM} system 2700 and the insertion site was determined by sequencing the PCR product (http://www.eurofinsdna.com).

Dissection and Sample Preparation

Flies were anaesthetized using CO_2, then immediately dissected. The tissues dissected were the complete heads, severed at the neck from adult flies and the CNS (developing brain), including the optic lobes from third instar wandering stage larvae. Tissues were collected into RLT buffer, pooled and extracted for RNA using Qiagen RNeasy RNA extraction kit (Qiagen, Valencia, CA, USA). In total there were 3 replicates for *Btk29A*^{ficP} and *Btk29A*$^{(fic\ Exc1-16)}$ sample groups (the larval CNS and adult heads).

RNA Isolation and Microarray Processing

RNA was extracted and *in vitro* reverse-transcribed according to Affymetrix protocol. Quality assurance was provided by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), after RNA extraction and *in vitro* transcription steps. *Drosophila* genome 2 expression arrays were hybridized and read using standard Affymetrix procedures. Microarrays were run at the Bionformatics and Expression Analysis core facility (http://apt.bea.ki.se/index.html) located at Karolinska Institutet, Huddinge (Novum).

Processing of the High-throughput Arrays and Analysis

GeneChip.CEL files were analyzed by using R statistical programming language, Bioconductor (http://www.bioconductor.org/), and the affy package. Data were initially RMA normalized first across the samples and then within each sample group. RMA normalized data were then scaled to a common median value. Both raw and pre-processed data is deposited in GEO (http://www.ncbi.nlm.nih.gov/geo/) (GSE30627).

Further filtering, within sample group analyses and pair-wise comparisons were carried out using dChip (https://sites.google.com/site/dchipsoft/, http://biosun1.harvard.edu/complab/dchip/) and Microsoft Excel.

For the Venn-diagram (Figure 7), the probe-sets found to be expressed above background (Affymetrix P-calls) in all three replicates per sample group were considered within this figure.

Comparison between Btk-dependent Transcripts from Btk KO Mouse Transitional Type 1 B-cells [40] and Adult Btk29A^{ficP} Drosophila Btk-dependent Transcripts

An inter-species comparison was conducted of the Btk-dependent transcripts identified in *Drosophila*, a total of 744 Btk-dependent transripts were found, compared to our previous transcriptomic data obtained from mouse Btk-defective Transitional type 1 B-lymphocytes [40], where a total of 147 Btk-dependent genes were reported as differentially expressed. The 147 genes found to be Btk-dependent in Btk-defective mouse Transitional type 1 B-cells were investigated for orthologs/homologs in the *Drosophila melanogaster* specie. We made use of the Affymetrix oligonucleotide array comparison to find which transcripts could be possible homologs (www.affymetrix.com/analysis/index.affx). The converted orthologs were then compared with the differentially expressed Btk-dependent transcripts found in the *Drosophila Btk29A*^{ficP} adult heads.

Gene Set Enrichment Analysis

Enriched GO clusters were analyzed using Cytoscape (http://www.cytoscape.org/) [26], with the plug-in system BiNGO [27] in addition to the DAVID web-tool (http://david.abcc.ncifcrf.gov/home.jsp) [28,29]. The Hyper-geometric Test with Benjamini-Hochberg False Discovery Rate Correction was chosen for both the analyses [27].

Supporting Information

Figure S1 Gene Set Enrichment Analysis for Gene Ontology (biological process) clusters for the 391 Btk-dependent transcripts during Fly neuronal development (Figure 4 in manuscript): Enriched Biological Process clusters within the list of 391-transcripts (Btk-dependent transcripts during fly neuronal development). The genes belonging to respective cluster are written next to the grey-zoned areas. For the main Figure 4, Figure S1 shows the respective genes found for each cluster (grey-zoned in Figure 4).
(EPS)

Figure S2 Gene Set Enrichment Analysis for Gene Ontology (biological process) clusters for the 167 Btk-dependent transcripts found in larvae CNS: Differentially expressed transcripts (167) were subjected to GSEA and enriched clusters were found. The genes belonging to respective cluster are marked in either red (up-regulated) or blue (down-regulated) depending on the direction of the gene. For the main Figure 5, Figure S2 shows the respective genes found for each cluster (grey-zoned in Figure 5). The genes belonging to respective cluster are marked in either red (up-regulated) or blue (down-regulated) depending on the direction of the gene.
(EPS)

Figure S3 Gene Set Enrichment Analysis for Gene Ontology (biological process) clusters for the 744 Btk-dependent transcripts found in adult head tissue: Enriched clusters of Biological Processes were found in the Btk-dependent Adult head tissue analysis and were statistically significant. The genes belonging to respective cluster are marked in either red (up-regulated) or blue (down-regulated) depending on the direction of the gene. For the main Figure 6, Figure S3 shows the respective genes found for each cluster (grey-zoned in Figure 6). The genes belonging to respective cluster are marked in either red (up-regulated) or blue (down-regulated) depending on the direction of the gene.
(EPS)

Table S1 Table S1 mirrors Table 1 (*25 genes found to overlap between D. melanogaster larvae CNS and adult head*) in addition of adding the dimension of every gene's Gene Ontology term including Gene Ontology ID.
(DOCX)

Author Contributions

Conceived and designed the experiments: JML CIES. Performed the experiments: HMN PK. Analyzed the data: JML HMN CIES. Contributed reagents/materials/analysis tools: NH DY. Wrote the paper: JML HMN CIES.

References

1. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188: 107–116.
2. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, et al. (2011) The evolution of gene expression levels in mammalian organs. Nature 478: 343–348.
3. Ortutay C, Nore BF, Vihinen M, Smith CI (2008) Phylogeny of Tec family kinases identification of a premetazoan origin of Btk, Bmx, Itk, Tec, Txk, and the Btk regulator SH3BP5. Adv Genet 64: 51–80.
4. Smith CI, Islam TC, Mattsson PT, Mohamed AJ, Nore BF, et al. (2001) The Tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, Tec, Txk and homologs in other species. Bioessays 23: 436–446.
5. Campana D, Farrant J, Inamdar N, Webster AD, Janossy G (1990) Phenotypic features and proliferative activity of B cell progenitors in X-linked agammaglobulinemia. J Immunol 145: 1675–1680.
6. Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, et al. (2002) Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. Pediatr Res 51: 159–168.
7. Monroe JG, Dorshkind K (2007) Fate decisions regulating bone marrow and peripheral B lymphocyte development. Adv Immunol 95: 1–50.
8. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, et al. (1993) Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell 72: 279–290.
9. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, et al. (1993) The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature 361: 226–233.
10. Rawlings DJ, Saffran DC, Tsukada S, Largaespada DA, Grimaldi JC, et al. (1993) Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. Science 261: 358–361.
11. Thomas JD, Sideras P, Smith CI, Vorechovsky I, Chapman V, et al. (1993) Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes. Science 261: 355–358.
12. Baba K, Takeshita A, Majima K, Ueda R, Kondo S, et al. (1999) The Drosophila Bruton's tyrosine kinase (Btk) homolog is required for adult survival and male genital formation. Mol Cell Biol 19: 4405–4413.
13. Hamada N, Backesjo CM, Smith CI, Yamamoto D (2005) Functional replacement of Drosophila Btk29A with human Btk in male genital development and survival. FEBS Lett 579: 4131–4137.
14. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715–720.
15. Guarnieri DJ, Dodson GS, Simon MA (1998) SRC64 regulates the localization of a Tec-family kinase required for Drosophila ring canal growth. Mol Cell 1: 831–840.
16. Roulier EM, Panzer S, Beckendorf SK (1998) The Tec29 tyrosine kinase is required during Drosophila embryogenesis and interacts with Src64 in ring canal development. Mol Cell 1: 819–829.
17. Chou TB, Perrimon N (1992) Use of a yeast site-specific recombinase to produce female germline chimeras in Drosophila. Genetics 131: 643–653.
18. Djagaeva I, Doronkin S, Beckendorf SK (2005) Src64 is involved in fusome development and karyosome formation during Drosophila oogenesis. Dev Biol 284: 143–156.
19. Siegfried E, Chou TB, Perrimon N (1992) wingless signaling acts through zeste-white 3, the Drosophila homolog of glycogen synthase kinase-3, to regulate engrailed and establish cell fate. Cell 71: 1167–1179.
20. Thomas JH, Wieschaus E (2004) src64 and tec29 are required for microfilament contraction during Drosophila cellularization. Development 131: 863–871.
21. Tateno M, Nishida Y, Adachi-Yamada T (2000) Regulation of JNK by Src during Drosophila development. Science 287: 324–327.
22. Lu N, Guarnieri DJ, Simon MA (2004) Localization of Tec29 to ring canals is mediated by Src64 and PtdIns(3,4,5)P3-dependent mechanisms. EMBO J 23: 1089–1100.
23. Chandrasekaran V, Beckendorf SK (2005) Tec29 controls actin remodeling and endoreplication during invagination of the Drosophila embryonic salivary glands. Development 132: 3515–3524.
24. Mukhopadhyay S, Ramars AS, Dash D (2001) Bruton's tyrosine kinase associates with the actin-based cytoskeleton in activated platelets. J Cell Biochem 81: 659–665.
25. Nore BF, Vargas L, Mohamed AJ, Branden LJ, Backesjo CM, et al. (2000) Redistribution of Bruton's tyrosine kinase by activation of phosphatidylinositol 3-kinase and Rho-family GTPases. Eur J Immunol 30: 145–154.
26. Sharma S, Orlowski G, Song W (2009) Btk regulates B cell receptor-mediated antigen processing and presentation by controlling actin cytoskeleton dynamics in B cells. J Immunol 182: 329–339.
27. Yao L, Janmey P, Frigeri LG, Han W, Fujita J, et al. (1999) Pleckstrin homology domains interact with filamentous actin. J Biol Chem 274: 19752–19761.
28. Liu C, Miller H, Hui KL, Grooman B, Bolland S, et al. (2011) A balance of Bruton's tyrosine kinase and SHIP activation regulates B cell receptor cluster formation by controlling actin remodeling. J Immunol 187: 230–239.
29. Dow JA, Romero MF (2010) Drosophila provides rapid modeling of renal development, function, and disease. Am J Physiol Renal Physiol 299: F1237–1244.
30. Read RD, Cavenee WK, Furnari FB, Thomas JB (2009) A drosophila model for EGFR-Ras and PI3K-dependent human glioma. PLoS Genet 5: e1000374.
31. Bina S, Wright VM, Fisher KH, Milo M, Zeidler MP (2010) Transcriptional targets of Drosophila JAK/STAT pathway signalling as effectors of haematopoietic tumour formation. EMBO Rep 11: 201–207.
32. Hinman VF, Nguyen AT, Cameron RA, Davidson EH (2003) Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci U S A 100: 13356–13361.
33. Angelini DR, Kaufman TC (2005) Insect appendages and comparative ontogenetics. Dev Biol 286: 57–77.
34. Erwin DH, Davidson EH (2009) The evolution of hierarchical gene regulatory networks. Nat Rev Genet 10: 141–148.
35. Hicks JK, Heitman J (2007) Divergence of protein kinase A catalytic subunits in Cryptococcus neoformans and Cryptococcus gattii illustrates evolutionary reconfiguration of a signaling cascade. Eukaryot Cell 6: 413–420.
36. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, et al. (2008) Rewiring the specificity of two-component signal transduction systems. Cell 133: 1043–1054.
37. Yeh BJ, Rutigliano RJ, Deb A, Bar-Sagi D, Lim WA (2007) Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. Nature 447: 596–600.
38. Islam TC, Branden LJ, Kohn DB, Islam KB, Smith CI (2000) BTK mediated apoptosis, a possible mechanism for failure to generate high titer retroviral producer clones. J Gene Med 2: 204–209.
39. Uckun FM (1998) Bruton's tyrosine kinase (BTK) as a dual-function regulator of apoptosis. Biochem Pharmacol 56: 683–691.
40. Lindvall JM, Blomberg KE, Berglof A, Smith CI (2006) Distinct gene expression signature in Btk-defective T1 B-cells. Biochem Biophys Res Commun 346: 461–469.
41. Lindvall JM, Blomberg KE, Berglof A, Yang Q, Smith CI, et al. (2004) Gene expression profile of B cells from Xid mice and Btk knockout mice. Eur J Immunol 34: 1981–1991.

Estimating the influence of $CD4^+$ $CD8^+$ T cell interactions on the transcriptome

Hossain M. Nawaz[1], K. Emelie M. Blomberg[1], Jessica M. Lindvall[2], C. I. Edvard Smith[1] and Anna Berglöf[1]

[1]Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Huddinge, Sweden

[2]Bioinformatics and Expression Analysis, Department of Biosciences and Nutrition, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Huddinge, Sweden

Corresponding authors:

Hossain M. Nawaz**,** Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Huddinge, Sweden. Tel: +46-858583657; Fax: +46-8-58583650; E-mail: hossain.nawaz@ki.se

And

C. I. Edvard Smith, Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Huddinge, Sweden. Tel: +46-858583651; Fax: +46-8-58583650; E-mail: edvard.smith@ki.se

Abstract

When evaluating the contribution of different T cell subsets for the immune response, many investigators first purify individual populations and subsequently activate them. In this study we have also followed this conventional method and compared gene expression profiling of mouse $CD3^+$, $CD4^+$ and $CD8^+$ T cells activated with anti-CD3 antibody. By using an algorithm-based analysis method, we found 6% of all expressed genes to be dependent on $CD4^+$-$CD8^+$ interactions. Gene set enrichment analysis (GSEA) of these genes revealed 22 overrepresented Gene Ontology (GO) Biological processes. Among those 'adaptive immune response' and 'lymphocyte/leucocyte mediated immunity' were of interest in the context of T cell activation. We also observed similar expression pattern in genes involved in immune response and T cell activation compared to the global expression profiling This observation has significant implications, especially when comparing gene expression data from different laboratories.

Introduction

Lymphocytes play a key role for the specific immune recognition of pathogens as they initiate adaptive immune responses. The interaction between B and T lymphocytes in immune responses began to be deciphered in the late 1960s and early 1970s[1]. Several *in vivo* and *in vitro* methods for obtaining relatively pure populations of T and B cells were also established[2]. Subpopulation of both B and T cells were successfully separated using different methods [3, 4]. *In vitro* stimulation of T lymphocytes has been frequently used to expand specific cells present at low precursor frequency *in vivo*[5]. However, this technique might not reflect the *in vivo* situation because of the variable efficiency of outgrowth of different T cell sub-populations [6]. Interaction of the $CD4^+/CD8^+$ lineages in the thymus is a complex process as both subsets arise from double-positive thymocytes [7]. Mature T cells thus have the surface phenotype $CD4^+/CD8^-$ or $CD4^-/CD8^+$ and their development is a highly regulated process[8, 9]. *In vitro* activation of both B and T cells has been performed on a regular basis in numerous research laboratories and one conventional way is to first separate the cells and activate them thereafter, using different stimuli [10-13]. Since the late 1990s, scientists have investigated changes in the percentage and state of activation of $CD4^+$ and $CD8^+$ T cells within efferent lymph draining antigen-stimulated peripheral lymphnodes and found that they have a greater percentage of $CD4^+$ cells and consequently a higher CD4 : CD8 ratio than peripheral blood [14, 15], whereas the percentage of $CD8^+$ cells is approximately equal in all four lymphocyte compartments in efferent lymph, afferent lymph, lymphoid tissue mononuclear cells and peripheral blood[16]. In the secondary lymphoid organ (SLO), the periarteriolar lymphoid sheath (PALS) of the spleen, the paracortex of the

lymph nodes and the interfollicular zone of Peyer's patches are collectively known as T cell zones [17]. They are densely packed with $CD4^+$ and $CD8^+$T cells and are sporadically scattered. Very little has been known so far whether these cells are in direct contact or not, but cytokines produced by these cells play a major role when it comes to $CD4^+$ and $CD8^+$ T cell interaction especially in virus infections [18, 19].

However, the influence of $CD4^+$-$CD8^+$ interaction for the establishment of the transcriptome still remains rather unexplored. In this study, using $CD3^+$, $CD4^+$ and $CD8^+$ T cells, we aimed to illustrate whether there is any difference if the cells are initially left un-separated *or* first separated, *and then* activated using existing expression data from previously conducted microarray studies. We performed an algorithm-based comparative analysis of the transcriptional pattern between two T cell subsets ($CD4^+$ and $CD8^+$) and the corresponding, un-separated, original $CD3^+$ population with a major focus on immune related genes using 'innateDB' Immune databases. Among others, we observed a significant influence on CD4-CD8 interaction for genes involved in immune response and T cell activation as measured at the mRNA level. We also identified differentially expressed genes after activation and changes in transcript levels using Gene Ontology (GO) analysis [20]. Gene set enrichment analysis (GSEA) [21] also revealed significant differences between the distribution of GO terms when the cells are separated first and then activated.

Materials and methods

T-cell separation and stimulation

$CD3^+$, $CD4^+$ and $CD8^+$ T-cells were isolated from pooled suspensions of spleen and lymph nodes of C57BL/6 wild type (Wt) mice using a negative selection method. [22]. The cells were separated over a MACS column (Miltenyi Biotec) and the cell suspensions were incubated with the antibodies in PBS supplemented with 2% FCS. Streptavidin beads (BD Pharmingen) were used for negative depletion according to manufacturer's instructions. The purity of the cells was assessed by flow cytometry and was routinely >90% $CD3^+$, >96% $CD4^+$ and >90% $CD8^+$ T-cells. T cells subsets ($CD3^+$, $CD4^+$ and $CD8^+$) were stimulated with anti-CD3 (1 μg/ml) for 24 h in 48-well plates. For each stimulus, at least duplicate samples were used in all but one experiment. For the $CD3^+$ T cells, triplicates and for the $CD4^+$ T cells duplicates were collected from Wt mice. For the $CD8^+$ T cells, we obtained a single sample from Wt owing to low yield of resting.

RNA isolation and microarray processing

Total RNA was extracted according to RNeasy Mini protocol (Qiagen, Valencia, CA, USA). The cRNA synthesis and hybridization were performed using Affymetrix MOE430 2.0 chips in BEA core facility at Department of Biosciences, Karolinska Institutet at Novum, Huddinge, Sweden. In total 10 arrays were analyzed which are accessible through the Gene Expression Omnibus (GEO; GSE12466).

Data analysis and bioinformatics approach

The pre-processing of the data was performed using MAS5 algorithm [23] in the Affymetrix Expression Array Console [24]. The pre-processed data were filtered upon the assessed background Signal criteria that were set to linear signal value of 50. Assessment of background noise was based upon the number of 'Absent call' in all the probe sets. Additional filtering was done after removing the Affymetrix control genes designated with suffixes –AFFX. In total 19,112 probe-sets fulfilled the criteria. Further analyses were carried out using Subio Platform (http://www.subio.jp/). Fold change criteria of $^{+}$/-2 was chosen for differentially expressed genes for each T cell subset.

A curated 'Immune Gene' list was obtained from the immunology database innateDB (http://www.innatedb.com/resourcesCurationGenes.jsp) where 1,402 innateDB Innate Immunity Genes (both human and mouse) have been annotated (dated March 2013) [25]. Among those, 483 mouse genes were used for the further analysis. Gene ontology (GO) analysis was carried out using *web*-based tool 'DAVID' (http://david.abcc.ncifcrf.gov/) [26]. Gene set enrichment analysis (GSEA) were exercised using Cytoscape with 'BiNGO' plug-in [27]. Filtering criteria was set for overrepresented GO terms 'biological process' with a False Discovery Rate (FDR) $p>0.05$ using Benjamini and Hochberg multiple testing correction [28].

Results

Global gene expression profiling of activated T cell subsets

Microarray gene expression analysis was carried out to observe the changes taking place at the transcriptomic level in T cells that were first separated into $CD4^+$ and $CD8^+$ population respectively and then either activated using anti- CD3 antibodies for 24 hours or left unstimulated for the same time period. We found that the unstimulated T cell subsets were clustered together (Figure 1) regardless if they were separated or not. This result implies that the effect of activation on the transcriptome is stronger than the differences among the T cell subsets themselves.

An amount of 3,299 transcripts were found to be differentially expressed in the $CD3^+$ mixed population after activation, with 1,870 transcripts being down-regulated and 1,429 transcripts found to be up-regulated. On the other hand, in separated $CD4^+$ T cells, after activation, 3,140 transcripts were found to be differentially expressed with a number of 1,590 down-regulated transcripts and 1,550 up-regulated transcripts. In $CD8^+$ T cells, 5,273 transcripts were found to differ with a fold-change of $^+$/- 2 and among those 3,172 transcripts were found to be down-regulated whereas, 2,101 transcripts were up-regulated. Thus, the pattern of differential gene expression is similar in all three T cell subsets, even if the predominance of down-regulated genes was not found in the $CD4^+$ population.

An amount of 887 transcripts were found to be ‘Common’ in all the T cell subsets after activation. Only 184 transcripts were found to be uniquely common between $CD3^+$ and

$CD4^+$ T cells, whereas the coressponding number was 1,499 transcripts overlapping between activated $CD3^+$ and $CD8^+$ and 991 transcripts between $CD4^+$ and $CD8^+$ T cells (Figure 2A). Thus, the relative overlap is similar between separated and unsepated cells.

Further analysis using GSEA shows the distribution of GO terms in T cell subsets (Figure 2B). 406 GO Biological Process (BP) terms resembling 887 transcripts were 'common' in all three subsets. In $CD3^+$ and $CD4^+$ T cells, 19 BP terms corresponding to 184 transcripts were found to be overlapping. This number was much higher in $CD3^+$ and $CD8^+$ (265) most probably due to high number of input transcripts (1,499). In $CD4^+$ and $CD8^+$ T cells, the number of common BP terms was 48. This distribution was in accordance to what we found in global expression profile. In depth analysis of the' common' 406 BP terms was carried out in order to see the expression patterns in the T cell subsets. GO BP terms were randomly selected to cover a more broad context, i.e. Metabolic process GO:0008152, Cell cycle GO:0007049, Apoptotic process GO:0006915, mRNA Transcription GO:0009299 and Translation GO:0006412. In 'mRNA Transcription', the ratio of up-and down-regulated genes was similar in $CD3^+$ and $CD4^+$subsets, whereas in $CD8^+$ T cells, the ratio was almost 2:1 (Figure 3 A). On the contrary, the number of down-regulated genes was very low when it comes to BP term 'Translation' but the pattern was similar in all three T cell subsets (Figure 3B).

The influence of $CD4^+$ and $CD8^+$ cell interactions

In order to find out the frequency of transcripts after activation of T cells and if the gene expression profile after stimulation was different when cells were cultured as un-separated ($CD4^+$ and $CD8^+$ together) compared to culture separately, we used a hypothesis-based algorithm on the expression data. Around 10,000 probe-sets containing 7,947 unique genes were selected with pre-defined criteria (see Materials and Methods). Signal intensities from $CD4^+$ or $CD8^+$ T cells after stimulation multiplied by the correction factor for frequency (= the percentage of $CD4^+$ and $CD8^+$ cells in the $CD3^+$ population (=1.7))[29]. Thus the correction factor was calculated as 0.63 for $CD4^+$ and 0.37 for $CD8^+$T cells.

*Signal intensity of each gene for $CD4^+$ * 0.63 = N, and*

*Signal intensity of each gene for $CD8^+$ * 0.37 = M.*

If we assume that the signal intensity (the values obtained after removal of background noise/signal <50, linear signal value) for gene 'X' in activated $CD3^+$, $CD4^+$ and $CD8^+$ T cell populations is Z, N and M respectively, then there are two mutually exclusive scenarios:

Scenario (A): If, $N+M = Z$ (Interpretation: no influence of one subset on another)

When, $\sum (N+M)/Z > 0.5$, *or* < 2

Or, *Scenario (B)*: If, $N+M \neq Z$ (Interpretation: there is an influence)

When, $\sum (N+M)/Z < 0.5$ *or* > 2

We found 7,474 (~94%) genes that fell into *Scenario A* (*no influence of one cell type on another*), while 462 (~6%) genes were observed in *Scenario B* (*there is an influence of one cell type on another*). While this demonstrates that for the majority of genes, separated versus co-cultivated $CD4^+$ and $CD8^+$ cells show the same result there is a population of genes where there is a significant difference. Gene set enrichment analysis (GSEA) for those 462 genes using GO Biological processes showed 22 significant BP terms (Table 1). Among those 'adaptive immune response' and 'lymphocyte/leucocyte mediated immunity' were of interest in the context of T cell activation.

Influence on immune response and T cell activation

In order to investigate 'Innate immunity genes' in T cell subsets upon activation, we obtained a list of 483 curated mouse genes in InnatedDB immunology database (http://www.innatedb.com/resourcesCurationGenes.jsp) . Out of these, 175 genes fell into the category of no influence pre-defined as *Scenario A*, whereas 18 genes were categorized as *Scenario B* according to our hypothesis-based algorithm. In a separate analysis, 89 genes were found to be differentially expressed in un-separated activated $CD3^+$ T cells in comparison to the resting state (Figure 4A). Among those, 25 genes were up-regulated whereas 64 genes were down-regulated. A similar kind of study in the $CD4^+$ and $CD8^+$ T cells populations showed 62 and 109 differentially expressed genes, respectively (supplementary table 1). The ratio of up-and down-regulated genes were not only similar in the separated T cell subsets but also resembled the global transcriptional

profiling as a majority of the genes were found to be down-regulated after activation. Using Gene Ontology GO, 31 genes were identified with the BP term ‘T cell activation’ in $CD3^+$ mixed population, whereas the corresponding number was 35 for the separated $CD4^+$ subset and 41 genes for the separated $CD4^+$ and $CD8^+$ T cell subset (Figure 4B and supplementary table 2).

Discussion

Genome-scale transcriptional profiling of T-cell activation has been studied with or without co-stimulation by anti-CD28 antibody [30-32]. A 'Medline' search dated July 2013 with search words 'Activation of $CD4^+$ and $CD8^+$ T cells after separation' revealed 128 articles where most of the research laboratories follow the more conventional method and separate lymphocytes before activation. In this study, we show that this may not be the optimal way since the separated cells show some altered gene expression profile**.** Global expression profiling in mouse T cell subsets showed discrepancy in the number of differentially expressed genes in activated $CD3^+$, $CD4^+$ and $CD8^+$ T cells respectively. We then exercised a hypothesis-based algorithm on the gene expression data, which unveils that 6% of all genes (462) are influenced if they are separated first and then activated. Thus, analysing data in two different approaches strengthens our supposition when it comes to the context of cultivating T cells *in vitro* whether it is separated first and then activated or the *vice versa*.

We conducted further down-stream analysis on those 462 genes using GSEA which revealed overrepresented 22 BP terms (Table 1). Among those, 'adaptive immune response' and 'lymphocyte mediated immunity' were two of the processes in the context of T cell activation and innate immunity as one of our prime interests was to find out if there is any altered expression pattern regarding immune response related genes in activated T cell subsets. Genes that were included in these biological processes - *Cd8a*,*Igha,Gzmb*,*Swap7*0,*H2-DMa*, and *Gadd45g*. *Cd8a* is known to be a cell surface

glycoprotein found on most cytotoxic T lymphocytes that mediates efficient cell-cell interactions within the immune system [33]. Granzyme B (*Gzmb*), which is expressed by cytotoxic T lymphocytes (CTL), is crucial for the rapid induction of target cell apoptosis by CTL in cell-mediated immune response [34]. Several well-known genes related with immune response and T cell activation like *Ifng, Cd44, Ilr4a* were also found related to various biological processes in this data-set. CD44 is known to be a marker for T cell development in the thymus and also an indicative marker for effector-memory T-cells [35]. Ilr4a is required for binding of interleukin-4 to the receptor alpha chain, which is a crucial event for the generation of a Th2-dominated early immune response [36]. IFNγ is the primary cytokine as Th1 cells secrete IFNγ, which in turn causes more undifferentiated $CD4^+$ cells (Th0 cells) to differentiate into Th1 cells and is critical for innate and adaptive immunity against viral and intracellular bacterial infections and for tumor control [37]. Thus, these well-known genes have influence when it comes to cell-cell interaction if the cells are co-cultivated and then stimulated. This observation might bring new insight in the context of *in vitro* T cell culture.

The roles of $CD4^+$ T cells in $CD8^+$ responses have been re-defined since the recovery of $CD8^+$ T_{reg} cells at the end of the primary response as $CD4^+$ help in the development of functional CD8 memory [38-40]. A recent comparative genome-scale, transcriptional profiling of T-cell activation in the $CD4^+$ and $CD8^+$subsets and the mixed $CD3^+$ populations also showed newly identified immune response genes in the context of communication between $CD4^+$ and $CD8^+$ T cells and T cell activation [41, 42]. We obtained a list of 483 curated mouse genes from the immune database innateDB and as

per our suggested algorithm, 91 % of these 483 genes that fulfils the selection criteria, according our pre-defined algorithm, fell into the category of 'no influence'. On the contrary, 9 % of these genes were considered to have an influence one cell to another ($CD4^+$- $CD8^+$). This finding suggests the role of cell-cell interaction in the context of innate immunity and to our knowledge it has not thoroughly been addressed in a systematic way.

There is no direct evidence regarding whether the $CD4^+$ and $CD8^+$ cells are in physical contact with each other. Studies on distribution of porcine T cell subsets showed similar number of $CD4^+$ and $CD8^+$T cells in the perifollicular region of the cortex but more $CD8^+$ T cells in medullary and cortico-medullary junctions in the lymphnodes. In spleen, $CD4^+$ T cells residing in the periarteriolar lymphoid sheath appeared to be more frequent than $CD8^+$ T cells [43]. On the other hand, significantly larger fraction of the $CD4^+$ lymphocyte population has been reported to be found in Peyer's patches and palatine tonsils than in lymph nodes [44]. Recent advances in two-photon microscopy has brought new insights in the context of the dynamics of immune-cell interactions in living tissues [45]. Distinct dynamics in the lymphnodes has been reported during the induction of antigen-specific immunity or tolerance using real-time two-photon microscopy [46]. Dendritic cells are professional antigen-presenting cells and also play an important role in T-cell peripheral tolerance with their unique property of inducing priming and differentiation of naïve $CD4^+$ and $CD8^+$ T cells into helper and cytotoxic effectors by secreting cytokines [47]. The role of $CD4^+$ T-cell in promoting primary $CD8^+$ T-cell responses as well as establishing functional memory $CD8^+$ T cells has also being shown

in various experimental settings [48]. The activated $CD8^+$ T cells showed poor survival without $CD4^+$ T cell help during both the primary and secondary responses [49].

GSEA showed 406 GO Biological Processes that were found to be common between $CD3^+$, $CD4^+$ and $CD8^+$ T cell subsets after activation. The BP term 'Translation' showed a temporal pattern as most of the genes were found to be up-regulated in all three subsets. This observation is supported by a previous work where Garcia-Sanz et al., showed translational repression during T cell activation [50]. Taken together, our study illustrates the interaction of $CD4^+$ and $CD8^+$ T cells, when they are purified or unseparated as $CD3^+$. Using expression profiling data, we found that 6% of the transcripts have an influence in the communication between $CD4^+$ and $CD8^+$ T cells. However our results suggest a small significance of in vitro cultivation of T cell subsets in the $CD4^+$-$CD8^+$ interaction. This finding is of general importance, since whenever T cells are cultivated, whether they are separated into subsets or not, has no such major influence on the induced transcriptomes.

Acknowledgements

The authors thank Akio Tanabe from Subio Platform for his bioinformatics help and expertise. We would also like to thank the Affymetrix Core facility at Novum. This work is supported by the Swedish Cancer Society and the Swedish Medical Council.

References

1. Miller JF, Mitchell GF: **Cell to cell interaction in the immune response**. *Transplant Proc* 1969, **1**(1):535-538.
2. Miller JF, Basten A, Sprent J, Cheers C: **Interaction between lymphocytes in immune responses**. *Cell Immunol* 1971, **2**(5):469-495.
3. Shortman K, Fidler JM, Schlegel RA, Nossal GJ, Howard M, Lipp J, von Boehmer H: **Subpopulations of B lymphocytes: physical separation of functionally distinct stages of B-cell differentiation**. *Contemp Top Immunobiol* 1976, **5**:1-45.
4. Shortman K, Von Boehmer H, Lipp J, Hopper K: **Subpopulations of T-lymphocytes. Physical separation, functional specialisation and differentiation pathways of sub-sets of thymocytes and thymus-dependent peripheral lymphocytes**. *Transplant Rev* 1975, **25**:163-210.
5. Burmester GR, Jahn B, Gramatzki M, Zacher J, Kalden JR: **Activated T cells in vivo and in vitro: divergence in expression of Tac and Ia antigens in the nonblastoid small T cells of inflammation and normal T cells activated in vitro**. *J Immunol* 1984, **133**(3):1230-1234.
6. Dietrich PY, Walker PR, Schnuriger V, Saas P, Perrin G, Guillard M, Gaudin C, Caignard A: **TCR analysis reveals significant repertoire selection during in vitro lymphocyte culture**. *Int Immunol* 1997, **9**(8):1073-1083.
7. Tanaka H, Taniuchi I: **The CD4/CD8 Lineages: Central Decisions and Peripheral Modifications for T Lymphocytes**. *Current topics in microbiology and immunology* 2013.
8. Naito T, Taniuchi I: **The network of transcription factors that underlie the CD4 versus CD8 lineage decision**. *International immunology* 2010, **22**(10):791-796.
9. Koretzky GA: **Multiple roles of CD4 and CD8 in T cell activation**. *J Immunol* 2010, **185**(5):2643-2644.
10. Hastings WD, Anderson DE, Kassam N, Koguchi K, Greenfield EA, Kent SC, Zheng XX, Strom TB, Hafler DA, Kuchroo VK: **TIM-3 is expressed on**

activated human CD4+ T cells and regulates Th1 and Th17 cytokines. *Eur J Immunol* 2009, **39**(9):2492-2501.

11. Armerding D, Sachs DH, Katz DH: **Activation of T and B lymphocytes in vitro. III. Presence of Ia determinants on allogeneic effect factor**. *J Exp Med* 1974, **140**(6):1717-1722.
12. Yamada-Ohnishi Y, Azuma H, Urushibara N, Yamaguchi M, Fujihara M, Kobata T, Ikeda H: **Cytotoxic difference of T cells expanded with anti-CD3 monoclonal antibody in the presence and absence of anti-CD 28 monoclonal antibody**. *Stem Cells Dev* 2004, **13**(3):315-322.
13. Tham EL, Mescher MF: **The poststimulation program of CD4 versus CD8 T cells (death versus activation-induced nonresponsiveness)**. *J Immunol* 2002, **169**(4):1822-1828.
14. Bujdoso R, Young P, Hopkins J, Allen D, McConnell I: **Non-random migration of CD4 and CD8 T cells: changes in the CD4: CD8 ratio and interleukin 2 responsiveness of efferent lymph cells following in vivo antigen challenge**. *European journal of immunology* 1989, **19**(10):1779-1784.
15. Abernethy NJ, Hay JB, Kimpton WG, Washington EA, Cahill RN: **Non-random recirculation of small, CD4+ and CD8+ T lymphocytes in sheep: evidence for lymphocyte subset-specific lymphocyte- endothelial cell recognition**. *International immunology* 1990, **2**(3):231-238.
16. Mackay CR, Kimpton WG, Brandon MR, Cahill RN: **Lymphocyte subsets show marked differences in their distribution between blood and the afferent and efferent lymph of peripheral lymph nodes**. *The Journal of experimental medicine* 1988, **167**(6):1755-1765.
17. Junt T, Scandella E, Ludewig B: **Form follows function: lymphoid tissue microarchitecture in antimicrobial immune defence**. *Nature reviews Immunology* 2008, **8**(10):764-775.
18. Su HC, Cousens LP, Fast LD, Slifka MK, Bungiro RD, Ahmed R, Biron CA: **CD4+ and CD8+ T cell interactions in IFN-gamma and IL-4 responses to viral infections: requirements for IL-2**. *J Immunol* 1998, **160**(10):5007-5017.
19. Pien GC, Nguyen KB, Malmgaard L, Satoskar AR, Biron CA: **A unique mechanism for innate cytokine promotion of T cell responses to viral infections**. *J Immunol* 2002, **169**(10):5827-5837.
20. Azuaje F, Al-Shahrour F, Dopazo J: **Ontology-driven approaches to analyzing data in functional genomics**. *Methods Mol Biol* 2006, **316**:67-86.
21. Tintle NL, Best AA, DeJongh M, Van Bruggen D, Heffron F, Porwollik S, Taylor RC: **Gene set analyses for interpreting microarray experiments on prokaryotic organisms**. *BMC bioinformatics* 2008, **9**:469.
22. Blomberg KE, Boucheron N, Lindvall JM, Yu L, Raberger J, Berglof A, Ellmeier W, Smith CE: **Transcriptional signatures of Itk-deficient CD3+, CD4+ and CD8+ T-cells**. *BMC genomics* 2009, **10**:233.
23. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**(2):185-193.
24. Auer H, Newsom DL, Kornacker K: **Expression Profiling Using Affymetrix GeneChip Microarrays**. *Methods Mol Biol* 2009, **509**:35-46.

25. Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan TH, Shah N *et al*: **InnateDB: facilitating systems-level analyses of the mammalian innate immune response**. *Molecular systems biology* 2008, **4**:218.
26. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic acids research* 2009, **37**(1):1-13.
27. Koh GC, Porras P, Aranda B, Hermjakob H, Orchard SE: **Analyzing protein-protein interaction networks**. *Journal of proteome research* 2012, **11**(4):2014-2031.
28. Hochberg Y, Benjamini Y: **More powerful procedures for multiple significance testing**. *Statistics in medicine* 1990, **9**(7):811-818.
29. Yoon HL, Singh KP, Ratner S, Reiners JJ, Jr.: **Phorbol ester effects on splenic lymphocyte composition and cytotoxic T cell activities of SSIN mice: a strain deficient in CD8+ T cells**. *Carcinogenesis* 1996, **17**(12):2617-2624.
30. Chtanova T, Newton R, Liu SM, Weininger L, Young TR, Silva DG, Bertoni F, Rinaldi A, Chappaz S, Sallusto F *et al*: **Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets**. *J Immunol* 2005, **175**(12):7837-7847.
31. Hughes-Fulford M, Sugano E, Schopper T, Li CF, Boonyaratanakornkit JB, Cogoli A: **Early immune response and regulation of IL-2 receptor subunits**. *Cellular signalling* 2005, **17**(9):1111-1124.
32. Diehn M, Alizadeh AA, Rando OJ, Liu CL, Stankunas K, Botstein D, Crabtree GR, Brown PO: **Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(18):11796-11801.
33. Doucey MA, Goffin L, Naeher D, Michielin O, Baumgartner P, Guillaume P, Palmer E, Luescher IF: **CD3 delta establishes a functional link between the T cell receptor and CD8**. *The Journal of biological chemistry* 2003, **278**(5):3257-3264.
34. Trapani JA, Sutton VR: **Granzyme B: pro-apoptotic, antiviral and antitumor functions**. *Current opinion in immunology* 2003, **15**(5):533-543.
35. Ponta H, Sherman L, Herrlich PA: **CD44: from adhesion molecules to signalling regulators**. *Nature reviews Molecular cell biology* 2003, **4**(1):33-45.
36. Kashiwada M, Giallourakis CC, Pan PY, Rothman PB: **Immunoreceptor tyrosine-based inhibitory motif of the IL-4 receptor associates with SH2-containing phosphatases and regulates IL-4-induced proliferation**. *J Immunol* 2001, **167**(11):6382-6387.
37. Schroder K, Hertzog PJ, Ravasi T, Hume DA: **Interferon-gamma: an overview of signals, mechanisms and functions**. *Journal of leukocyte biology* 2004, **75**(2):163-189.
38. Bourgeois C, Tanchot C: **Mini-review CD4 T cells are required for CD8 T cell memory generation**. *European journal of immunology* 2003, **33**(12):3225-3231.
39. Bourgeois C, Veiga-Fernandes H, Joret AM, Rocha B, Tanchot C: **CD8 lethargy in the absence of CD4 help**. *European journal of immunology* 2002, **32**(8):2199-2207.

40. Shedlock DJ, Shen H: **Requirement for CD4 T cell help in generating functional CD8 T cell memory**. *Science* 2003, **300**(5617):337-339.
41. Wang M, Windgassen D, Papoutsakis ET: **Comparative analysis of transcriptional profiling of CD3+, CD4+ and CD8+ T cells identifies novel immune response players in T-cell activation**. *BMC Genomics* 2008, **9**:225.
42. Wang M, Windgassen D, Papoutsakis ET: **A global transcriptional view of apoptosis in human T-cell activation**. *BMC medical genomics* 2008, **1**:53.
43. Jonjic N, Jonjic S, Saalmuller A, Rukavina D, Koszinowski UH: **Distribution of T-lymphocyte subsets in porcine lymphoid tissues**. *Immunology* 1987, **60**(3):395-401.
44. Zuckermann FA, Gaskins HR: **Distribution of porcine CD4/CD8 double-positive T lymphocytes in mucosa-associated lymphoid tissues**. *Immunology* 1996, **87**(3):493-499.
45. Cahalan MD, Parker I, Wei SH, Miller MJ: **Two-photon tissue imaging: seeing the immune system in a fresh light**. *Nature reviews Immunology* 2002, **2**(11):872-880.
46. Hugues S, Fetler L, Bonifaz L, Helft J, Amblard F, Amigorena S: **Distinct T cell dynamics in lymph nodes during the induction of tolerance and immunity**. *Nature immunology* 2004, **5**(12):1235-1242.
47. Hivroz C, Chemin K, Tourret M, Bohineust A: **Crosstalk between T lymphocytes and dendritic cells**. *Critical reviews in immunology* 2012, **32**(2):139-155.
48. Wiesel M, Oxenius A: **From crucial to negligible: functional CD8(+) T-cell responses and their dependence on CD4(+) T-cell help**. *European journal of immunology* 2012, **42**(5):1080-1088.
49. Novy P, Quigley M, Huang X, Yang Y: **CD4 T cells are required for CD8 T cell survival during both primary and memory recall responses**. *J Immunol* 2007, **179**(12):8243-8251.
50. Garcia-Sanz JA, Mikulits W, Livingstone A, Lefkovits I, Mullner EW: **Translational control: a general mechanism for gene regulation during T cell activation**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 1998, **12**(3):299-306.

Figure legends

Figure 1: Hierarchical clustering showing distribution of T cell subsets among the sample group using Subio platform. Activated T cells subsets are clustered together irrespective of separation. (Colour code showing the log intensity of the signal value, red=high intensity, blue=low intensity).

Figure 2: Venn diagram A. Number of overlapping transcripts in differentially expressed lists in different T cell subsets after activation. B. Number of overlapping GO terms in Gene set enrichment analysis (GSEA).

Figure 3: A. Bar graph showing alteration in gene expression for GO term 'mRNA Transcription derived from Gene set enrichment analysis (GSEA) and B. in 'Translation'.

Figure 4: A. Bar graph showing number of Immune gene in each T cell subsets after activation; B. Diagram showing number of genes from the differentially expressed list in T cell subsets assigned by GO term 'T cell Activation' using DAVID gene ontology tool.

Table 1: List of overrepresented GO Biological processes with corrected *P-value* in *scenario B* (*there is an influence of one cell type on another*) along with corresponding genes in each group. (Red denotes up-regulation and blue denotes down-regulation respectively)

Figure 1

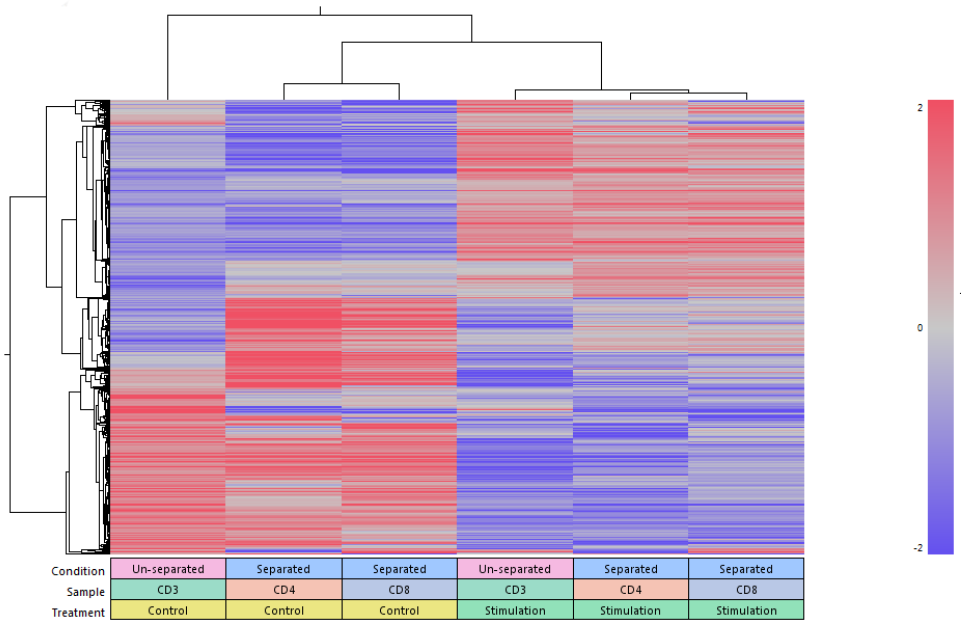
2
0
-2
Condition
Sample
Treatment
Un-separated
Separated
Separated
Un-separated
Separated
Separated
CD3
CD4
CD8
CD3
CD4
CD8
Control
Control
Control
Stimulation
Stimulation
Stimulation

Figure 2A

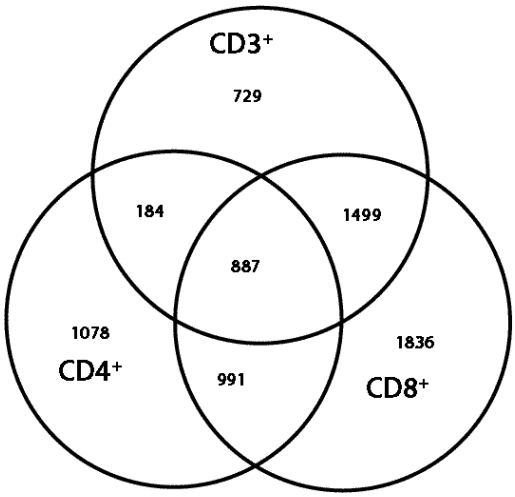

Figure 2B

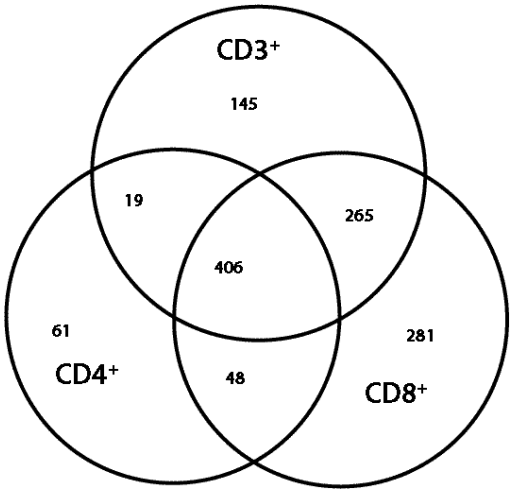

Figure 3 A

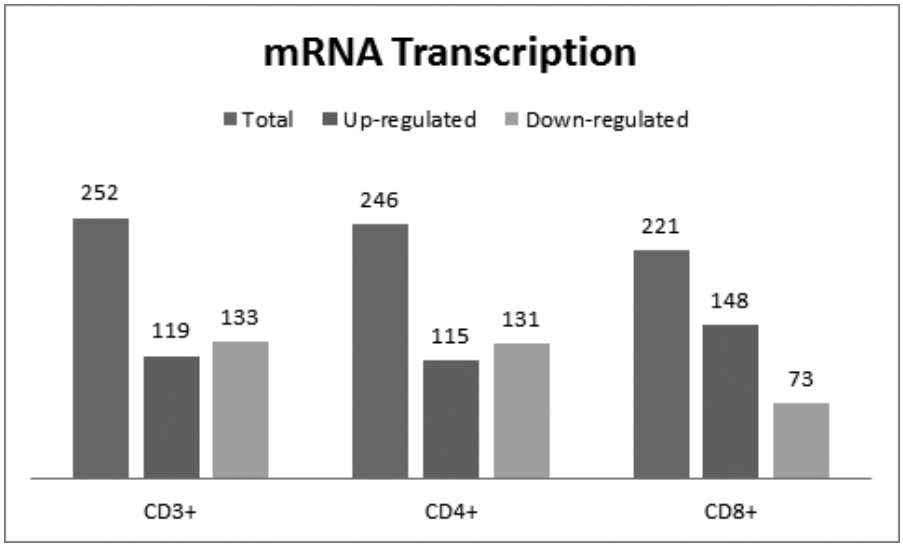

Figure 3 B

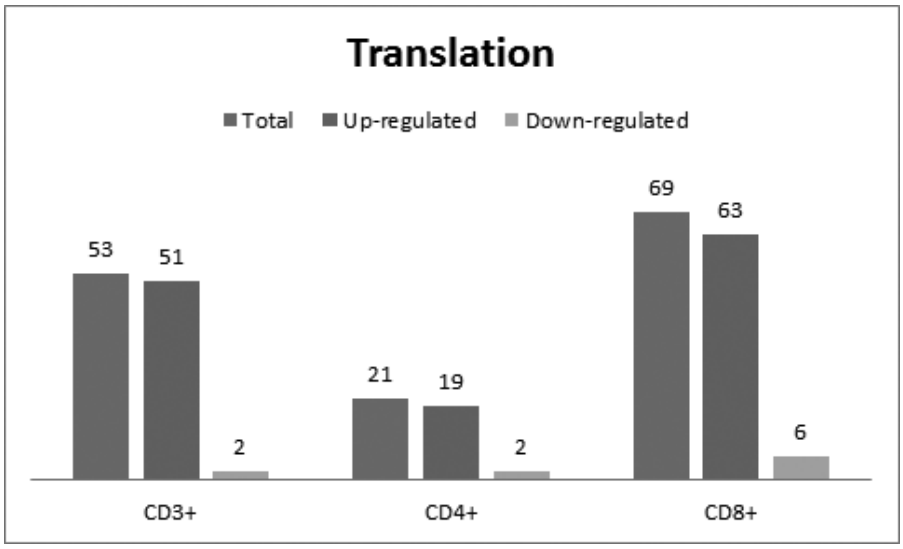

Figure 4 A

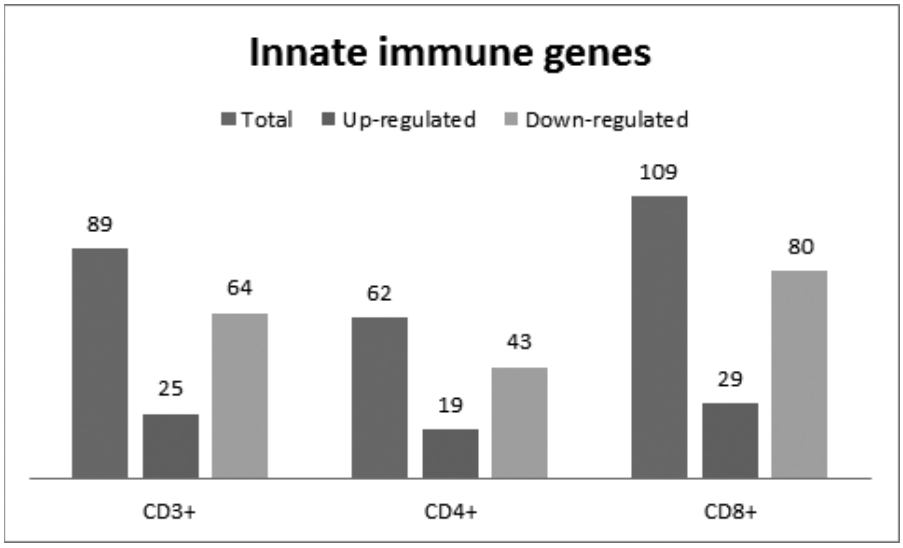

Figure 4 B

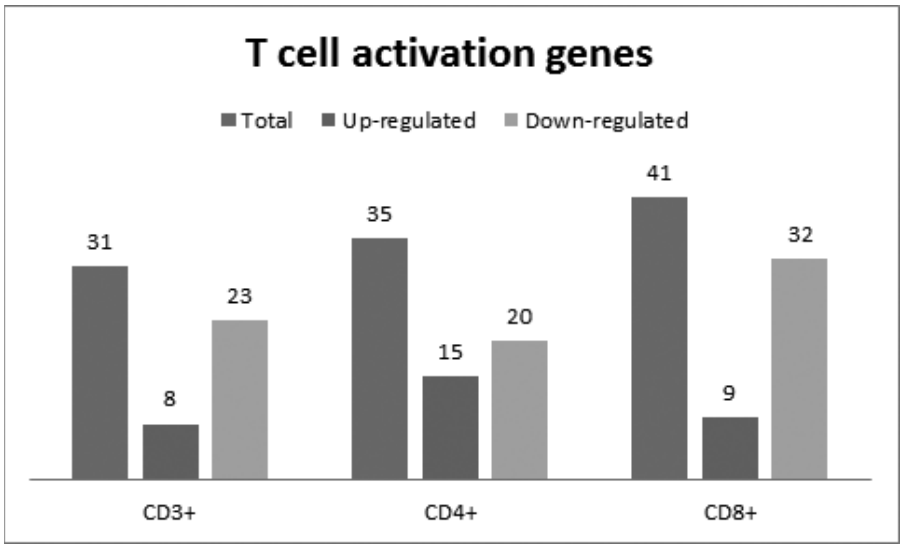

Table 1:

GO-ID	Description	Corrected P-value	Gene symbol
2250	adaptive immune response	0,0038721	*Cd8a* *Gzmb* *Swap70* *H2-DMa* *Gadd45g*
2460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0,0038721	*Cd8a* *Gzmb* *Swap70* *H2-DMa* *Gadd45g*
19882	antigen processing and presentation	0,010835	*H2-Aa* *Ifng* *H2-DMa* *H2-Eb1*
2478	antigen processing and presentation of exogenous peptide antigen	0,03683	*H2-Aa* *Ifng* *H2-DMa*
19886	antigen processing and presentation of exogenous peptide antigen via MHC class II	0,015294	*H2-Aa* *Ifng* *H2-DMa*
2495	antigen processing and presentation of peptide antigen via MHC class II	0,015294	*H2-Aa* *Ifng* *H2-DMa*
2504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0,024212	*H2-Aa* *Ifng* *H2-DMa*
30330	DNA damage response, signal transduction by p53 class mediator	0,015294	*Ifi204* *Trp53* *Bbc3*
42771	DNA damage response, signal transduction by p53	0,0091253	*Ifi204* *Trp53*

	class mediator resulting in induction of apoptosis		*Bbc3*
8630	DNA damage response, signal transduction resulting in induction of apoptosis	0,03683	*Ifi204* *Trp53* *Bbc3*
2443	leukocyte mediated immunity	0,025433	*Cd8a* *Gzmb* *Swap70* *H2-DMa*
2449	lymphocyte mediated immunity	0,010846	*Cd8a* *Gzmb* *Swap70* *H2-DMa*
35264	multicellular organism growth	0,035873	*Ankrd11* *Trp53* *Zfx* *Tnks2*
1779	natural killer cell differentiation	0,026922	*Sp3* *Ikzf1*
46638	positive regulation of alpha-beta T cell differentiation	0,034677	*Cd83* *Ikzf1* *Il4ra**
33138	positive regulation of peptidyl-serine phosphorylation	0,020704	*Ifng* *Akt2* *Cd44*
46637	regulation of alpha-beta T cell differentiation	0,043596	*Cd83* *Ikzf1* *Il4ra**
1558	regulation of cell growth	0,0044062	*Mndal,* *Apbb2* *Trp53* *Htra3*** *Cd44* *Ppan* *Bbc3*
33135	regulation of peptidyl-serine phosphorylation	0,040549	*Ifng* *Akt2*

			Cd44
28	ribosomal small subunit assembly	0,043824	*Rps10* *Rps15*
42274	ribosomal small subunit biogenesis	0,0056314	*Nob1* *Rps10* *Rps15*
2246	wound healing involved in inflammatory response	0,0054254	*Cd44* *Hif1a*

*Up-regulated in $CD3^+$ T cell. ** Up-regulated in $CD8^+$ T cell.

Supplementary Table 1: Differentially expressed lists of genes in 'Innate immune gene' in $CD3^+$ (A), $CD4^+$ (B) and $CD8^+$ (C). (The most 10 up-and down-regulated genes are marked with grey shades)

A:

Gene symbol	Entrez Gene	Gene Title	**Fold Change**
Pros1	19128	protein S (alpha)	**13.72431**
Nr4a3	18124	nuclear receptor subfamily 4, group A, member 3	**9.137889**
Irf8	15900	interferon regulatory factor 8	**7.9287004**
Ddx21	56200	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	**5.1723394**
Tnfrsf9	21942	tumor necrosis factor receptor superfamily, member 9	**4.8939753**
Cdk6	12571	cyclin-dependent kinase 6	**4.7233734**
Serpinb9	20723	serine (or cysteine) peptidase inhibitor, clade B, member 9	**4.4513044**
Pdcd1	18566	programmed cell death 1	**3.5218158**
Ddx1	104721	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	**3.4440618**
Ubqln1	56085	ubiquilin 1	**3.398982**
Eif4e	13684	eukaryotic translation initiation factor 4E	**3.289648**
Cops5	26754	COP9 (constitutive photomorphogenic) homolog, subunit 5 (Arabidopsis thaliana)	**2.9387605**
Tbx21	57765	T-box 21	**2.857089**
Hspa14	50497	heat shock protein 14	**2.7411203**
Rxra	20181	retinoid X receptor alpha	**2.7168**
Hspd1	15510	heat shock protein 1 (chaperonin)	**2.623722**
Scarb1	20778	scavenger receptor class B, member 1	**2.6210725**
Tpst1	22021	protein-tyrosine sulfotransferase 1	**2.4418108**
Hif1a	15251	hypoxia inducible factor 1, alpha subunit	**2.3304186**
Cflar	12633	CASP8 and FADD-like apoptosis regulator	**2.2872581**
Gbp2	14469	guanylate binding protein 2	**2.2051938**
Tufm	233870	Tu translation elongation factor, mitochondrial	**2.1401458**
Map3k7	26409	mitogen-activated protein kinase kinase kinase 7	**2.0636647**
Atf4	11911	activating transcription factor 4	**2.058319**
Aim2	383619	absent in melanoma 2	**0.4985011**
Cbl	12402	Casitas B-lineage lymphoma	**0.49138477**
Unc5cl	76589	unc-5 homolog C (C. elegans)-like	**0.48212025**
Zc3hav1	78781	zinc finger CCCH type, antiviral 1	**0.47195724**
Nfatc3	18021	nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 3	**0.46377388**
Pik3c3	225326	phosphoinositide-3-kinase, class 3	**0.45709905**
Glrx	93692	glutaredoxin	**0.44666728**
Tnfrsf1a	21937	tumor necrosis factor receptor superfamily, member 1a	**0.4403006**
Samhd1	56045	SAM domain and HD domain, 1	**0.4386051**
Rac2	19354	RAS-related C3 botulinum substrate 2	**0.4376938**

Foxo3	56484	forkhead box O3	**0.4281101**
S100a10	20194	S100 calcium binding protein A10 (calpactin)	**0.42044848**
Ing4	28019	inhibitor of growth family, member 4	**0.41375417**
Nfe2l2	18024	nuclear factor, erythroid derived 2, like 2	**0.39289626**
Pycard	66824	PYD and CARD domain containing	**0.38949293**
Nr3c1	14815	nuclear receptor subfamily 3, group C, member 1	**0.37565327**
Stat1	20846	signal transducer and activator of transcription 1	**0.36403802**
Elf1	13709	E74-like factor 1	**0.34310332**
Inpp5d	16331	inositol polyphosphate-5-phosphatase D	**0.3393214**
Atg12	67526	autophagy related 12	**0.33638194**
Ifih1	71586	interferon induced with helicase C domain 1	**0.3333476**
Grn	14824	granulin	**0.33177298**
Trim21	20821	tripartite motif-containing 21	**0.32586756**
Zbp1	58203	Z-DNA binding protein 1	**0.3198898**
Treml2	328833	triggering receptor expressed on myeloid cells-like 2	**0.31267655**
Pklr	18770	pyruvate kinase liver and red blood cell	**0.30733898**
Arhgap15	76117	Rho GTPase activating protein 15	**0.30332914**
Ifit2	15958	interferon-induced protein with tetratricopeptide repeats 2	**0.29759368**
Pla2g4a	18783	phospholipase A2, group IVA (cytosolic, calcium-dependent)	**0.29557285**
Aqp3	11828	aquaporin 3	**0.28769308**
Rora	19883	RAR-related orphan receptor alpha	**0.28195822**
Tsc22d3	14605	TSC22 domain family, member 3	**0.2817741**
Ctss	13040	cathepsin S	**0.2667402**
Dusp1	19252	dual specificity phosphatase 1	**0.26421064**
Anxa2	12306	annexin A2	**0.25611457**
Cd97	26364	CD97 antigen	**0.25553998**
Akna	100182	AT-hook transcription factor	**0.24614528**
Naip2	17948	NLR family, apoptosis inhibitory protein 2	**0.24460384**
Il1rl2	107527	interleukin 1 receptor-like 2	**0.23999679**
Cxcr3	12766	chemokine (C-X-C motif) receptor 3	**0.23075046**
Ikbke	56489	inhibitor of kappaB kinase epsilon	**0.22613524**
Itpr3	16440	inositol 1,4,5-triphosphate receptor 3	**0.21983248**
Rnasel	24014	ribonuclease L (2', 5'-oligoisoadenylate synthetase-dependent)	**0.21919127**
Ern1	78943	endoplasmic reticulum (ER) to nucleus signalling 1	**0.2149797**
Oas2	246728	2'-5' oligoadenylate synthetase 2	**0.21007602**
Btk	12229	Bruton agammaglobulinemia tyrosine kinase	**0.20821184**
Trib2	217410	tribbles homolog 2 (Drosophila)	**0.20576853**
Plec	18810	plectin	**0.19965765**
Oas1c	114643	2'-5' oligoadenylate synthetase 1C	**0.1971484**
Cyld	74256	cylindromatosis (turban tumor syndrome)	**0.19113925**
Nod1	107607	nucleotide-binding oligomerization domain containing 1	**0.17759846**

Oas1a	246730	2'-5' oligoadenylate synthetase 1A	**0.16102402**
Ctsb	13030	cathepsin B	**0.15992896**
Abcg1	11307	ATP-binding cassette, sub-family G (WHITE), member 1	**0.15162231**
Tnfaip8l2	69769	tumor necrosis factor, alpha-induced protein 8-like 2	**0.145865**
Tlr1	21897	toll-like receptor 1	**0.11134235**
Ddx58	230073	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	**0.10940291**
Tecpr1	70381	tectonin beta-propeller repeat containing 1	**0.1079904**
Ifit3	15959	interferon-induced protein with tetratricopeptide repeats 3	**0.10682104**
Txnip	56338	thioredoxin interacting protein	**0.09463819**
Pyhin1	236312	pyrin and HIN domain family, member 1	**0.08606364**
Pglyrp2	57757	peptidoglycan recognition protein 2	**0.0810056**
Abca1	11303	ATP-binding cassette, sub-family A (ABC1), member 1	**0.06897664**
Ifit1	15957	interferon-induced protein with tetratricopeptide repeats 1	**0.06884942**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.061535567**

B:

Gene Symbol	Entrez Gene	Gene Title	**Fold Change**
Pros1	19128	protein S (alpha)	**12.3801365**
Tnfrsf9	21942	tumor necrosis factor receptor superfamily, member 9	**7.183595**
Plscr1	22038	phospholipid scramblase 1	**5.723464**
Scarb1	20778	scavenger receptor class B, member 1	**5.40047**
E2f1	13555	E2F transcription factor 1	**4.2441173**
Hif1a	15251	hypoxia inducible factor 1, alpha subunit	**4.0453978**
Rxra	20181	retinoid X receptor alpha	**2.900475**
Traf6	22034	TNF receptor-associated factor 6	**2.8435144**
Nampt	59027	nicotinamide phosphoribosyltransferase	**2.6458852**
Lgals3	16854	lectin, galactose binding, soluble 3	**2.4340785**
Pmaip1	58801	phorbol-12-myristate-13-acetate-induced protein 1	**2.275942**
Ptges2	96979	prostaglandin E synthase 2	**2.2385283**
Trp53	22059	transformation related protein 53	**2.2198038**
Ddx21	56200	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	**2.1691952**
Map3k7	26409	mitogen-activated protein kinase kinase kinase 7	**2.1666358**
Tnf	21926	tumor necrosis factor	**2.1663227**
Hspa14	50497	heat shock protein 14	**2.1332688**
Bid	12122	BH3 interacting domain death agonist	**2.0124643**
Bcl3	12051	B cell leukemia/lymphoma 3	**0.49710596**
Ly96	17087	lymphocyte antigen 96	**0.4940673**
Cflar	12633	CASP8 and FADD-like apoptosis regulator	**0.4875447**
Zbp1	58203	Z-DNA binding protein 1	**0.484787**

Nr3c1	14815	nuclear receptor subfamily 3, group C, member 1	**0.4804168**
Ifnar1	15975	interferon (alpha and beta) receptor 1	**0.47693443**
Unc5cl	76589	unc-5 homolog C (C. elegans)-like	**0.47221813**
Ctsb	13030	cathepsin B	**0.4586592**
Treml2	328833	triggering receptor expressed on myeloid cells-like 2	**0.43873534**
Itch	16396	itchy, E3 ubiquitin protein ligase	**0.43145657**
Zc3hav1	78781	zinc finger CCCH type, antiviral 1	**0.4311032**
Cltc	67300	clathrin, heavy polypeptide (Hc)	**0.40575516**
Dok3	27261	docking protein 3	**0.40555874**
Ahr	11622	aryl-hydrocarbon receptor	**0.39366692**
Coro2a	107684	coronin, actin binding protein 2A	**0.39264557**
Il1rl2	107527	interleukin 1 receptor-like 2	**0.38478047**
Il1rl1	17082	interleukin 1 receptor-like 1	**0.3832328**
Pltp	18830	phospholipid transfer protein	**0.3790714**
Oas3	246727	2'-5' oligoadenylate synthetase 3	**0.37848514**
Aqp3	11828	aquaporin 3	**0.34386227**
Cyld	74256	cylindromatosis (turban tumor syndrome)	**0.3404207**
Oas1c	114643	2'-5' oligoadenylate synthetase 1C	**0.33922407**
Chuk	12675	conserved helix-loop-helix ubiquitous kinase	**0.3367694**
Naip2	17948	NLR family, apoptosis inhibitory protein 2	**0.3356082**
Rnasel	24014	ribonuclease L (2', 5'-oligoisoadenylate synthetase-dependent)	**0.29989034**
Rictor	78757	RPTOR independent companion of MTOR, complex 2	**0.28426898**
Mapk14	26416	Mitogen-activated protein kinase 14	**0.2787412**
Rora	19883	RAR-related orphan receptor alpha	**0.27558288**
Nod1	107607	nucleotide-binding oligomerization domain containing 1	**0.2572262**
Plec	18810	plectin	**0.25403547**
Rictor	78757	RPTOR independent companion of MTOR, complex 2	**0.25264958**
Trib2	217410	tribbles homolog 2 (Drosophila)	**0.25035408**
Rora	19883	RAR-related orphan receptor alpha	**0.24753278**
Ifit3	15959	interferon-induced protein with tetratricopeptide repeats 3	**0.22749074**
Ccr6	12458	chemokine (C-C motif) receptor 6	**0.1986626**
Akna	100182	AT-hook transcription factor	**0.19638774**
Ikbke	56489	inhibitor of kappaB kinase epsilon	**0.18553475**
Irf3	54131	interferon regulatory factor 3	**0.1833945**
Rnasel	24014	ribonuclease L (2', 5'-oligoisoadenylate synthetase-dependent)	**0.16128139**
Ddx58	230073	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	**0.1545684**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.1009478**
Irf7	54123	interferon regulatory factor 7	**0.09965712**
Oas1b	23961	2'-5' oligoadenylate synthetase 1B	**0.08353805**

C:

Gene Symbol	Entrez Gene	Gene Title	**Fold Change**
Pdcd1	18566	programmed cell death 1	**16.239805**
Nr4a3	18124	nuclear receptor subfamily 4, group A, member 3	**12.760379**
Tnf	21926	tumor necrosis factor	**7.7738776**
Ddx21	56200	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	**4.4549575**
Tpst1	22021	protein-tyrosine sulfotransferase 1	**4.4405**
Tnfrsf9	21942	tumor necrosis factor receptor superfamily, member 9	**34.302475**
Tnfsf11	21943	tumor necrosis factor (ligand) superfamily, member 11	**31.460342**
Gbp2	14469	guanylate binding protein 2	**3.663119**
Serpinb9	20723	serine (or cysteine) peptidase inhibitor, clade B, member 9	**3.581412**
Pik3ap1	83490	phosphoinositide-3-kinase adaptor protein 1	**3.5144367**
Ddx1	104721	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	**3.4350262**
Plscr1	22038	phospholipid scramblase 1	**3.4169834**
Nampt	59027	nicotinamide phosphoribosyltransferase	**3.1733992**
Irf8	15900	interferon regulatory factor 8	**28.209078**
Tbx21	57765	T-box 21	**2.7896063**
Foxo3	56484	forkhead box O3	**2.7538667**
Ubqln1	56085	ubiquilin 1	**2.6693857**
Scarb1	20778	scavenger receptor class B, member 1	**2.617003**
Cops5	26754	COP9 (constitutive photomorphogenic) homolog, subunit 5 (Arabidopsis thaliana)	**2.5603132**
Cflar	12633	CASP8 and FADD-like apoptosis regulator	**2.464877**
Eif4e	13684	eukaryotic translation initiation factor 4E	**2.4226782**
Hspa14	50497	heat shock protein 14	**2.3176851**
Hspd1	15510	heat shock protein 1 (chaperonin)	**2.250849**
Tufm	233870	Tu translation elongation factor, mitochondrial	**2.2464745**
Map3k7	26409	mitogen-activated protein kinase kinase kinase 7	**2.2161098**
Casp7	12369	caspase 7	**2.2058234**
March5	69104	membrane-associated ring finger (C3HC4) 5	**2.0846524**
Il1b	16176	interleukin 1 beta	**2.0652885**
Traf6	22034	TNF receptor-associated factor 6	**2.0505247**
Bcl2	12043	B cell leukemia/lymphoma 2	**0.49794412**
Dok3	27261	docking protein 3	**0.49775374**
Usp4	22258	ubiquitin specific peptidase 4 (proto-oncogene)	**0.4959911**
Pacsin1	23969	protein kinase C and casein kinase substrate in neurons 1	**0.4954406**
Il1rl2	107527	interleukin 1 receptor-like 2	**0.48886737**
Nrip1	268903	nuclear receptor interacting protein 1	**0.48659888**
Casp8	12370	caspase 8	**0.48656544**
Ifit2	15958	interferon-induced protein with tetratricopeptide repeats 2	**0.48382193**
Irak4	266632	interleukin-1 receptor-associated kinase 4	**0.46958405**

Nfatc3	18021	nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 3	**0.464011**
Eif2ak2	19106	eukaryotic translation initiation factor 2-alpha kinase 2	**0.46299917**
Lgals9	16859	lectin, galactose binding, soluble 9	**0.46189544**
Nod1	107607	nucleotide-binding oligomerization domain containing 1	**0.45942372**
Trim21	20821	tripartite motif-containing 21	**0.45827174**
Akt1	11651	thymoma viral proto-oncogene 1	**0.4526772**
Tax1bp1	52440	Tax1 (human T cell leukemia virus type I) binding protein 1	**0.4464951**
Ikbkb	16150	inhibitor of kappaB kinase beta	**0.44294488**
Nr3c1	14815	nuclear receptor subfamily 3, group C, member 1	**0.44061738**
Zc3hav1	78781	zinc finger CCCH type, antiviral 1	**0.44029713**
Cbl	12402	Casitas B-lineage lymphoma	**0.4384928**
Nfe2l2	18024	nuclear factor, erythroid derived 2, like 2	**0.43645692**
Pura	19290	purine rich element binding protein A	**0.4364379**
Adar	56417	Adenosine deaminase, RNA-specific	**0.43621206**
Cd97	26364	CD97 antigen	**0.4340961**
Arhgap15	76117	Rho GTPase activating protein 15	**0.4304531**
Mapk14	26416	Mitogen-activated protein kinase 14	**0.41837218**
Btk	12229	Bruton agammaglobulinemia tyrosine kinase	**0.41789111**
Cd300lf	246746	CD300 antigen like family member F	**0.41435316**
Inpp5d	16331	inositol polyphosphate-5-phosphatase D	**0.41123167**
Ing4	28019	inhibitor of growth family, member 4	**0.40929478**
Dusp10	63953	dual specificity phosphatase 10	**0.4069025**
Ifnar1	15975	interferon (alpha and beta) receptor 1	**0.40334955**
Itch	16396	itchy, E3 ubiquitin protein ligase	**0.40263137**
Elf1	13709	E74-like factor 1	**0.39930362**
Zc3h12a	230738	zinc finger CCCH type containing 12A	**0.39838165**
Ikbke	56489	inhibitor of kappaB kinase epsilon	**0.394835**
Ctsb	13030	cathepsin B	**0.38660493**
Ltbr	17000	lymphotoxin B receptor	**0.38316503**
Lair1	52855	leukocyte-associated Ig-like receptor 1	**0.37261263**
Ip6k1	27399	inositol hexaphosphate kinase 1	**0.37246168**
S100a10	20194	S100 calcium binding protein A10 (calpactin)	**0.37204677**
Akap10	56697	A kinase (PRKA) anchor protein 10	**0.36266223**
Naip2	17948	NLR family, apoptosis inhibitory protein 2	**0.3561497**
Aim2	383619	absent in melanoma 2	**0.3514144**
Pyhin1	236312	pyrin and HIN domain family, member 1	**0.3502625**
Irf3	54131	interferon regulatory factor 3	**0.3484335**
Cyld	74256	cylindromatosis (turban tumor syndrome)	**0.34648833**
Gsk3b	56637	Glycogen synthase kinase 3 beta	**0.3425342**
Akna	100182	AT-hook transcription factor	**0.3414414**
Ccr6	12458	chemokine (C-C motif) receptor 6	**0.322682**

Stim1	20866	stromal interaction molecule 1	**0.3209075**
Cnot4	53621	CCR4-NOT transcription complex, subunit 4	**0.3161724**
Aqp3	11828	aquaporin 3	**0.3159503**
Rictor	78757	RPTOR independent companion of MTOR, complex 2	**0.30611745**
Rora	19883	RAR-related orphan receptor alpha	**0.2977626**
Itpr3	16440	inositol 1,4,5-triphosphate receptor 3	**0.2975865**
Rnasel	24014	ribonuclease L (2', 5'-oligoisoadenylate synthetase-dependent)	**0.29466242**
Oas1c	114643	2'-5' oligoadenylate synthetase 1C	**0.2943554**
Oas1b	23961	2'-5' oligoadenylate synthetase 1B	**0.2733997**
Oas2	246728	2'-5' oligoadenylate synthetase 2	**0.25806507**
Ern1	78943	endoplasmic reticulum (ER) to nucleus signalling 1	**0.2514128**
Abcg1	11307	ATP-binding cassette, sub-family G (WHITE), member 1	**0.24762993**
Tecpr1	70381	tectonin beta-propeller repeat containing 1	**0.24462765**
Ifit3	15959	interferon-induced protein with tetratricopeptide repeats 3	**0.23268622**
Pglyrp2	57757	peptidoglycan recognition protein 2	**0.23127589**
Trib2	217410	tribbles homolog 2 (Drosophila)	**0.22072867**
Plec	18810	plectin	**0.21930467**
Ddx58	230073	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	**0.20914766**
Tnfaip8l2	69769	tumor necrosis factor, alpha-induced protein 8-like 2	**0.19629578**
Txnip	56338	thioredoxin interacting protein	**0.19268131**
Tlr1	21897	toll-like receptor 1	**0.17818478**
Nlrc3	268857	NLR family, CARD domain containing 3	**0.15994185**
Oas1a	246730	2'-5' oligoadenylate synthetase 1A	**0.15550193**
Oas3	246727	2'-5' oligoadenylate synthetase 3	**0.14724338**
Treml2	328833	triggering receptor expressed on myeloid cells-like 2	**0.1412012**
Ifit1	15957	interferon-induced protein with tetratricopeptide repeats 1	**0.13018186**
Abca1	11303	ATP-binding cassette, sub-family A (ABC1), member 1	**0.12118901**
Irf7	54123	interferon regulatory factor 7	**0.11746269**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.09431074**
Tsc22d3	14605	TSC22 domain family, member 3	**0.06060853**

Supplementary Table 2: Differentially expressed lists of genes in 'T cell activation' in $CD3^+$ (A), $CD4^+$ (B) and $CD8^+$ (C). (The most 10 up-and down-regulated genes are marked with grey shades)

A:

Gene Symbol	Entrez Gene	Gene Title	**Fold Change**
Zbtb32	58206	zinc finger and BTB domain containing 32	**5.1060586**
Gadd45g	23882	growth arrest and DNA-damage-inducible 45 gamma	**4.538945**
Stat5a	20850	signal transducer and activator of transcription 5A	**3.6064506**
Hspd1	15510	heat shock protein 1 (chaperonin)	**2.623722**
1443703_at			**2.5738592**
Jmjd6	107817	jumonji domain containing 6	**2.4585385**
Fkbp1a	14225	FK506 binding protein 1a	**2.1731253**
Blm	12144	Bloom syndrome, RecQ helicase-like	**2.1514957**
Ppp3cb	19056	protein phosphatase 3, catalytic subunit, beta isoform	**0.48641378**
Stat5b	20851	signal transducer and activator of transcription 5B	**0.47643435**
Ncor1	20185	nuclear receptor co-repressor 1	**0.44225666**
Myh9	17886	myosin, heavy polypeptide 9, non-muscle	**0.41647974**
Satb1	20230	special AT-rich sequence binding protein 1	**0.40193048**
Sox4	20677	SRY-box containing gene 4	**0.38886943**
Rpl22	100504863 /// 19934	60S ribosomal protein L22-like /// ribosomal protein L22	**0.37483567**
Smad3	17127	SMAD family member 3	**0.37025616**
Dock2	94176	dedicator of cyto-kinesis 2	**0.32776883**
H2-M3	14991	histocompatibility 2, M region locus 3	**0.31174248**
Hsh2d	209488	Hematopoietic SH2 domain containing	**0.29436725**
Itpkb	320404	inositol 1,4,5-trisphosphate 3-kinase B	**0.24133293**
Themis	210757	thymocyte selection associated	**0.23927073**
Lck	16818	lymphocyte protein tyrosine kinase	**0.20723325**
Rhoh	74734	ras homolog gene family, member H	**0.19496359**
Ccnd3	12445	cyclin D3	**0.17938279**
Sla2	77799	Src-like-adaptor 2	**0.17053519**
Card11	108723	caspase recruitment domain family, member 11	**0.1485169**
Cd86	12524	CD86 antigen	**0.1371939**
5830411N06 Rik	244234	RIKEN cDNA 5830411N06 gene	**0.1005823**
Il7r	16197	interleukin 7 receptor	**0.0751177**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.061535567**
Cd1d1	12479	CD1d1 antigen	**0.055453822**

B:

Gene Symbol	Entrez Gene	Gene Title	**Fold Change**
Slamf1	27218	signaling lymphocytic activation molecule family member 1	**8.749248**
1434279_at			**6.5629764**
Blm	12144	Bloom syndrome, RecQ helicase-like	**5.676049**
Gadd45g	23882	growth arrest and DNA-damage-inducible 45 gamma	**4.3227243**
Rbpj	19664	recombination signal binding protein for immunoglobulin kappa J region	**4.1750455**
Eomes	13813	eomesodermin homolog (Xenopus laevis)	**3.5600965**
Zbtb32	58206	zinc finger and BTB domain containing 32	**3.2836645**
Exo1	26909	exonuclease 1	**21.687248**
Stat5a	20850	signal transducer and activator of transcription 5A	**2.9683588**
Traf6	22034	TNF receptor-associated factor 6	**2.8435144**
Tpd52	21985	tumor protein D52	**2.6236224**
Pou2f2	18987	POU domain, class 2, transcription factor 2	**2.3865778**
Ercc1	13870	excision repair cross-complementing rodent repair deficiency, complementation group 1	**2.3435528**
Trp53	22059	transformation related protein 53	**2.2198038**
Dnaja3	83945	DnaJ (Hsp40) homolog, subfamily A, member 3	**2.0760431**
Bcl6	12053	B cell leukemia/lymphoma 6	**0.49855477**
Bcl3	12051	B cell leukemia/lymphoma 3	**0.49710596**
Ifnar1	15975	interferon (alpha and beta) receptor 1	**0.47693443**
1442617_at			**0.4578828**
Unc13d	70450	unc-13 homolog D (C. elegans)	**0.44653553**
Pik3cd	18707	phosphatidylinositol 3-kinase catalytic delta polypeptide	**0.434327**
Mll5	69188	myeloid/lymphoid or mixed-lineage leukemia 5	**0.38439572**
Clcf1	56708	cardiotrophin-like cytokine factor 1	**0.32848728**
Sla2	77799	Src-like-adaptor 2	**0.32388505**
Skap2	54353	src family associated phosphoprotein 2	**0.31942636**
Cd86	12524	CD86 antigen	**0.2755583**
Cxcr5	12145	chemokine (C-X-C motif) receptor 5	**0.2720764**
Lax1	240754	lymphocyte transmembrane adaptor 1	**0.26196453**
Pik3r1	18708	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)	**0.19818771**
Sbno2	216161	strawberry notch homolog 2 (Drosophila)	**0.18271326**
Themis	210757	thymocyte selection associated	**0.17829123**
5830411N06Rik	244234	RIKEN cDNA 5830411N06 gene	**0.16991177**
Ccnd3	12445	cyclin D3	**0.1490088**
Lck	16818	lymphocyte protein tyrosine kinase	**0.13601607**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.1009478**

C:

Gene Symbol	Entrez Gene	Gene Title	**Fold Change**
Blm	12144	Bloom syndrome, RecQ helicase-like	**4.551019**
Gadd45g	23882	growth arrest and DNA-damage-inducible 45 gamma	**4.341037**
Cd28	12487	CD28 antigen	**3.589423**
Fkbp1a	14225	FK506 binding protein 1a	**3.0457816**
Psen1	19164	presenilin 1	**2.3035903**
Hspd1	15510	heat shock protein 1 (chaperonin)	**2.250849**
Dnaja3	83945	DnaJ (Hsp40) homolog, subfamily A, member 3	**2.1950712**
Jmjd6	107817	jumonji domain containing 6	**2.1394832**
Zbtb32	58206	zinc finger and BTB domain containing 32	**19.175894**
Bcl2	12043	B cell leukemia/lymphoma 2	**0.49794412**
Apc	11789	adenomatosis polyposis coli	**0.49776313**
Atp7a	11977	ATPase, Cu++ transporting, alpha polypeptide	**0.49527687**
Prkdc	19090	protein kinase, DNA activated, catalytic polypeptide	**0.48708066**
Dock2	94176	dedicator of cyto-kinesis 2	**0.48291415**
Bcl11b	58208	B cell leukemia/lymphoma 11B	**0.47466266**
Smad3	17127	SMAD family member 3	**0.47357482**
Ncor1	20185	nuclear receptor co-repressor 1	**0.47129598**
Ptprc	19264	Protein tyrosine phosphatase, receptor type, C	**0.46320382**
Ppp3cb	19056	protein phosphatase 3, catalytic subunit, beta isoform	**0.4417437**
Stat5b	20851	signal transducer and activator of transcription 5B	**0.4247084**
H2-DMa	14998	histocompatibility 2, class II, locus DMa	**0.4131365**
Ifnar1	15975	interferon (alpha and beta) receptor 1	**0.40334955**
Myh9	17886	myosin, heavy polypeptide 9, non-muscle	**0.4006346**
Sox4	20677	SRY-box containing gene 4	**0.38668674**
Themis	210757	thymocyte selection associated	**0.38636786**
Gm15472	100503170	predicted gene 15472	**0.38031948**
LOC100504863 /// Rpl22	100504863 /// 19934	60S ribosomal protein L22-like /// ribosomal protein L22	**0.37140003**
Hsh2d	209488	Hematopoietic SH2 domain containing	**0.34332898**
Rhoh	74734	ras homolog gene family, member H	**0.29548007**
Card11	108723	caspase recruitment domain family, member 11	**0.27250442**
Itpkb	320404	inositol 1,4,5-trisphosphate 3-kinase B	**0.26633596**
Ccnd3	12445	cyclin D3	**0.22779824**
Cd1d1	12479	CD1d1 antigen	**0.17777559**
Nlrc3	268857	NLR family, CARD domain containing 3	**0.15994185**
Cd86	12524	CD86 antigen	**0.15750249**
Sla2	77799	Src-like-adaptor 2	**0.12787358**
Lck	16818	lymphocyte protein tyrosine kinase	**0.11883396**

Satb1	20230	special AT-rich sequence binding protein 1	**0.09969395**
P2rx7	18439	purinergic receptor P2X, ligand-gated ion channel, 7	**0.09431074**
5830411N06Rik	244234	RIKEN cDNA 5830411N06 gene	**0.0869558**
Il7r	16197	interleukin 7 receptor	**0.08614151**